languages

Special Issue Reprint

Syntax and Discourse at the Crossroads

Edited by
Ana Ojea

mdpi.com/journal/languages

MDPI

Syntax and Discourse at the Crossroads

Syntax and Discourse at the Crossroads

Guest Editor

Ana Ojea

MDPI

Basel • Beijing • Wuhan • Barcelona • Belgrade • Novi Sad • Cluj • Manchester

Guest Editor
Ana Ojea
Department of English,
French and German Philology
Oviedo University
Oviedo
Spain

Editorial Office
MDPI AG
Grosspeteranlage 5
4052 Basel, Switzerland

This is a reprint of the Special Issue, published open access by the journal *Languages* (ISSN 2226-471X), freely accessible at: www.mdpi.com/journal/languages/special_issues/9B887SQ07R.

For citation purposes, cite each article independently as indicated on the article page online and using the guide below:

Lastname, Firstname, Firstname Lastname, and Firstname Lastname. Article Title. *Journal Name* **Year**, *Volume Number*, Page Range.

ISBN 978-3-7258-3594-2 (Hbk)
ISBN 978-3-7258-3593-5 (PDF)
https://doi.org/10.3390/books978-3-7258-3593-5

Contents

About the Editor

Ana Ojea

Ana Ojea is an Associate Professor of Linguistics in the Department of English, French and German Philology at Oviedo University (Spain). She specializes in syntax from a generativist perspective, with a particular emphasis on English and Spanish, and the topics she has addressed include the analysis of relative sentences, verbal modification, defective clauses, and the syntactic expression of modality; she has also explored the process of first language acquisition from both theoretical and empirical perspectives. Currently, she focuses on the syntax–pragmatics interface, with medium-term objectives that include distinguishing the informational features that are part of the speakers' grammatical competence from those that are exclusively linked to their pragmatic competence.

languages

MDPI

Editorial

Introduction to the Special Issue *Syntax and Discourse at the Crossroads*

Ana Ojea [1,2]

[1] Department of English, French and German Philology, University of Oviedo, 33011 Oviedo, Asturias, Spain; aojea@uniovi.es
[2] Departamento de Filología Inglesa, Francesa y Alemana, Facultad de Filosofía y Letras, Universidad de Oviedo, 33011 Oviedo, Asturias, Spain

1. This Special Issue

This Issue features eleven papers that explore significant aspects of the syntax–discourse interface using empirical data from different languages (English, German, Spanish, French, and Italian) and their variants. The majority of these papers adopt a formal generative approach and attempt to demarcate the object of study of formal linguistics in addressing the core question of the limits of syntax—that is, whether syntax is self-contained with respect to pragmatics or whether notions of discourse should be included in the syntactic derivation instead, as in the so-called syntactization of discourse. I will briefly discuss here three details of the relationship between syntax and pragmatics that have attracted scholarly attention in recent generative studies and then summarize the papers in this issue, highlighting their contributions to this ongoing debate. The final section outlines the potential paths they pave for further research in this field.

2. The Syntax–Discourse Interface

Given the importance of external interfaces with syntax in the Minimalist Program, the connection between syntax and pragmatics, which is programmatic in the functional theories of language, is studied extensively in present-day generative grammar. In his most recent work, Chomsky has argued that the only uniquely human component of the faculty of language (the so-called *faculty of language in the narrow sense*) essentially consists of a computational mechanism, *Merge*, that is capable of generating recursive structures. According to this view, the remaining properties of language originate from the interaction of this mechanism with other mental systems not exclusively devoted to language, such as the articulatory perceptual system (i.e., what we can hear and say, or see and sign, and our capacity to process sounds) and the conceptual–intentional system (i.e., the need to successfully convey meaning and pragmatic intentions); for further reading on this, see, among others, Hauser et al. (2002); Pinker and Jackendoff (2005); and Mendívil-Giró (2018) and the references therein. Consequently, we are spurred to ask which pragmatic information conditions sentence form and should therefore be syntactically encoded in terms of specific categories and discourse features.

On the basis of seminal work by Rizzi (1997), a general consensus has emerged that at least the different types of topics and foci must be syntactically represented, along with the types of information that they convey, essentially correlated with what is taken to be given and new information in the sentence. Researchers therefore seek to provide a fine-grained analysis of the left periphery of the clause, defining the sets of hierarchically ordered projections and features that articulate the different types of topicalized and/or focalized

Received: 28 January 2025
Revised: 11 February 2025
Accepted: 13 February 2025
Published: 25 February 2025

Citation: Ojea, A. (2025). Introduction to the Special Issue *Syntax and Discourse at the Crossroads*. *Languages*, *10*(3), 36. https://doi.org/10.3390/languages10030036

constituents displaced to the periphery of the sentence under particular communicative conditions (see Rizzi, 1997; Cinque, 1999; Frascarelli & Hinterhölzl, 2007; Cruschina, 2012, a.o.).

Another fruitful area of research has focused on the relationship between syntax and the speech act and how it affects the structure of the left periphery. This line of inquiry emphasizes the necessity of incorporating discourse participants such as the speaker and the addressee, along with concepts like commitment, evidentiality, evaluation, mirativity, point of view or sentience which are pertinent to them. These investigations have provided principled explanations for various phenomena whose grammatical properties are substantially influenced by the discursive status of the proposition (see Ross, 1970; Speas & Tenny, 2003; Haegeman, 2014; Miyagawa, 2022; Krifka, 2023, a.o.).

Finally, there is also a growing body of literature on the syntactic, semantic, and discursive properties of the subject. The subject is the DP (a verbal argument or an expletive) that marks the morphological agreement with the verb, and languages have traditionally been parametrized in terms of their SV/VS order and whether null subjects can be licensed in them or not. Indeed, many generative works have discussed the structural position of the subject and the licensing requirements for null categories in its position, together with the discursive import of this lexical/null DP in a sentence. This is crucially connected to the opposition between categorical statements, where something is asserted about an entity, and thetic statements, which are logically unstructured and recognize a certain state of affairs in a given place or at a given time (see, a.o., Kuroda, 1972 and Sasse, 1987); terminological pairs such as declarative/presentational or sentence-focus/predicate-focus sentences refer to this same dichotomy. In categorical statements, the subject is generally understood to be the subject of predication, that is, the aboutness topic of the sentence signaling what the sentence is about (the intentional base, in Ojea, 2017's terms); in thetic statements, however, the subject is simply an entity involved in the eventuality expressed rather than the topic itself.

Significantly, the categorical/thetic dichotomy finds some structural correlation in Romance languages, with categorical sentences unmarkedly displaying an SV order and thetic sentences a VS order. In relation, some authors contend that the interpretative properties of preverbal subjects in categorical statements can only result from their placement in a non-argumental position within the CP-layer, while others argue that preverbal subjects should sit in Spec-TP, the canonical subject position crosslinguistically, irrespectively of the intentional reading of the sentence (see Lobo & Martins, 2017 for an overview of this controversy). Another point of contention is whether the propositional content of all clauses should be checked against a topic, which would imply that a topic–comment articulation applies to thetic statements too and that they are predicated of a stage topic, as suggested by Erteschik-Shir (1997).

The eleven papers in this volume, which have been arranged alphabetically, serve to delineate the intersection of syntax and discourse further in their presentation of innovative research on grammatical descriptions, the linguistic variation at both the micro and macro levels, code-switching scenarios, and diachronic evolution. These studies also lay the groundwork for future advances in the field since they address languages and constructions understudied so far and adopt methodological tools that facilitate the collection of large samples of naturalistic data.

3. Summary of the Contributions to This Special Issue

Delia Bentley and **Francesco Maria Ciconte's** contribution to this Special Issue is situated within the framework of Role and Reference Grammar, a parallel architecture theory which represents discourse–pragmatics and semantics separately from syntax. Their

paper focuses on presentational sentences, which, as argued above, are characterized by a verb–subject (VS) order in Romance languages. In certain dialects in Northern Italy, these constructions sometimes incorporate an etymologically locative presentational clitic. By employing data gathered through questionnaire-assisted interviews conducted in situ, Bentley and Ciconte investigate the patterns of the microvariation in these presentational clitics in Milanese and Turinese. Their findings indicate that in Milanese, the clitic *ghe*, which originated as a resumptive locative pronoun, has evolved into a subject agreement marker. However, in Turinese, alternative findings are uncovered regarding the clitic *je*, as this clitic does not preclude the number agreement between the verb and the postverbal noun phrase. These authors argue that *je* in Turinese is not the syntactic subject but rather can be interpreted as the subject of predication, serving as a marker of the aboutness topic of the utterance. The analysis in their paper therefore hinges on a critical distinction between the concepts of the subject of predication and the syntactic subject, which, as previously noted, need not align.

In their study of the behavior of the expletives *a* and *chiru* in Fornese and Cilentano, two understudied Romance varieties spoken in Northern and Southern Italy, respectively, **Simone De Cia** and **Mariangela Cerullo** also use corpus data to analyze linguistic variation. Utilizing novel data collected during several field trips, they show that *a* and *chiru* occur in the same syntactic contexts as subject expletives in non-pro-drop languages but argue that they actually are discourse-pragmatic expletives and as such, the manifestation a formal requirement at the syntax-pragmatics interface. This requirement forces Fornese and Cilentano to satisfy the discourse feature [aboutness] structurally, hence why the expletives *a* and *chiru* are externally merged as a last-resort strategy to indicate the lack of aboutness, i.e., the absence of an aboutness referent in the utterance, as applies in thetic statements. Nevertheless, the lexicalization of the discourse–pragmatic expletive does not target the same functional projection in these two languages but is subject to parametric variation: *a* in Fornese occupies the higher portion of the TP-domain, whereas *chiru* in Cilentano satisfies [aboutness] in the canonical syntactic position for overt aboutness-shift topics within the C-domain.

Mara Frascarelli and **Giorgio Carella**'s paper tackles the occurrence of null subjects in non-pro-drop languages, highlighting the intricate interplay between the structural, semantic, and discourse factors behind the omission of the subject. It analyzes data drawn from three transcripts from two online corpora of Modern Colloquial French, namely the CLAPI and the CFPP2000. Their statistical analysis of the results reveals that the overwhelming majority of null subjects in Colloquial French result from the omission of expletive subjects in predicational sentences. Thus, Frascarelli and Carella contend that the subject *il* in these predicational contexts is an expletive pronoun and susceptible to being dropped; in contrast, *ce*, the subject of non-predicational copular sentences in French, is not a true expletive but rather a referential pronoun, making its omission by speakers less likely. The authors also demonstrate that referential DPs may be omitted when they establish an Agree relationship with a referent that is strongly active in the current discourse context, i.e., an aboutness-shift topic. Lastly, in comparing French and English, they confirm that the omission of the subject is permissible in both languages but that referential subjects are omitted significantly more frequently in English than they are in French.

In his investigation of the positioning of subjects, **Roland Hinterhölzl** correlates the distribution of DPs in German with the phenomenon of complementizer agreement in West Germanic languages. He assumes that the context values become accessible in a specific functional head in the C-domain and proposes that in German, indefinite DPs and weak quantifiers are interpreted within the V-domain, while attributively used definite DPs and strong quantifiers (unless they are discourse-anaphoric) are interpreted in the T-domain.

Conversely, referentially used definite DPs and anaphoric strong quantifiers (i.e., discourse-anaphoric DPs) must access the C-domain and enter into a licensing relationship with the head Fin^0 to be fully licensed. Hinterhölzl then outlines the pattern of subject licensing in Cimbrian, a German dialect spoken in the village of Luserna, Trentino, to illustrate how this pattern is replicated in the systems of complementizer agreement (CA) in West Germanic dialects, a multifaceted and complex phenomenon that is more substantial than a formal peculiarity. He demonstrates that the data from Cimbrian and the CA in West Germanic show striking similarities and warrant a unified explanation in which CA is understood as serving the purpose of anchoring the subject into the context.

Ángel L. Jiménez-Fernández and **Mercedes Tubino-Blanco** discuss inferential interrogative sentences with *qué* in Spanish, which exhibit the shape of a *wh*-question but the interpretation of a polar question and convey a reading of evidentiality (e.g., *¿Qué vas, en coche?* 'Are you going by car (I infer)?'). They propose a multi-layer system with two contrasted levels in their internal composition—an utterance level, configured as an interrogative speech act, and a clausal level—and discuss the prominent roles of the speaker and the addressee in the derivation. Although the evidential element in inferential interrogative sentences has an unusual speaker-oriented interpretation, Jiménez-Fernández and Tubino-Blanco demonstrate that the Interrogative Flip—a noteworthy cross-linguistic feature of interrogatives with morphologically overt evidentials—is also activated in these constructions due to the way in which the speaker and addressee are anchored within the discourse: the speaker draws the inference, while the addressee is solicited for confirmation. Their formal analysis therefore accounts for this unconventional interpretation of the interrogative clause, its hybrid nature as both a *wh-* and *yes/no* question, and its distinctive intonation pattern of fall–rise.

The paper by **Sergio López-Martínez** examines Old English, which has traditionally been characterized as a V2 language. In particular, it analyzes a productive presentational construction from this period in which an adverb or a prepositional phrase indicating location occupies the leftmost position of the clause, resulting in inversion of the finite verb. For López-Martínez, this construction is equivalent to locative inversion in Present-Day English, which has often been viewed as a root phenomenon. However, as demonstrated in this article, it can occur quite productively in non-root subordinate clauses in Old English. Utilizing data from the York–Toronto–Helsinki Parsed Corpus of Old English Prose, López-Martínez qualitatively and quantitively analyzes various types of subordinate constructions that feature clause-initial prepositional phrases (PPs) with and without finite verb inversion. He documents a relatively balanced distribution of PP-S-V and PP-V-S word orders in subordinate sentences and investigates the factors that contribute to the occurrence of a PP-V-S order in these clauses in Old English. He concludes that while the type of verb is a critical determinant of the inversion in main clauses, in subordinate clauses, the length and type of the subject are more critical, with longer and heavier subjects triggering finite verb inversion more readily.

María Mare focuses on the so-called $PRO_{[PL]}$ *with*-DP construction in examples such as "*Con mi mujer nos casamos en abril*", "My wife and I got married in April", which can only be fully understood by involving syntactic and informational considerations. The verb in this type of construction, as observed in some varieties of Spanish (such as Argentinian and Chilean Spanish), exhibits dual number information despite the fact that the reference of the argument, introduced by *con,* "with", is a single entity. Additionally, both holistic and distributive interpretations of the predicate are permitted, even if a comitative item that generally voids a distributive reading is present. Mare reexamines this kind of construction and proposes that the *with*-DP phrase constitutes part of a complex subject DP containing an empty plural pronominal, which may be doubled by an overt pronoun being positioned at

the left periphery of the sentence. She approaches the analysis from an information structure perspective, with two mechanisms determining the referential properties of the pronominal and the plural number agreement with the verb: one involving the features associated with person-related information, such as the combination of [−Participant]/[+Author], and the other pertaining to distinctions related to anaphoricity.

Research addressing the intricate relationship between code-switching—the integration of multiple languages into a single sentence or discourse—and the organizational principles that govern information structure is currently limited. However, **Antje Muntendam** and **M. Carmen Parafita Couto's** paper aims to address this gap by providing a comprehensive overview of the current, albeit sparse, research concerning the intersection between code-switching and information packaging. It is widely understood that the information structure across languages can be conveyed through various means, including syntax (e.g., word order), morphology (e.g., the use of topic and focus markers), and prosody (e.g., intonation). Consequently, code-switching between languages that employ distinct methods for expressing information structure is expected. To illustrate this phenomenon, Muntendam and Parafita Couto examine specific instances of code-switching at the interfaces, demonstrating how the study of interfaces informs the study of code-switching (and vice versa) and allows hypotheses to be tested that cannot be examined using monolingual data alone. They also acknowledge the limitations of the existing research and discuss certain theoretical and methodological considerations that should guide future studies on code-switching, emphasizing the need to use naturalistic data to properly analyze the prosody, syntax, and morphology involved in the expression of the information structure within code-switched speech.

My own paper, "The Syntax of Speech Acts: Deictic Inversion as an Evidential Strategy in English", offers a novel analysis of deictic inversion (DI) in English, a construction which is used by a speaker to gesture towards a proximal or distal location, directing the addressee's attention to an entity related to said location (e.g., "*Here comes the bus")*. Superficially, DI has much in common with standard locative inversion (LI): In both cases, the sentence is understood as a non-predicative assertion of a state of affairs where the grammatical subject receives the informational focus. Furthermore, these two constructions are headed by the same types of verbs (the copula *be* or unaccusative predicates) and feature a locative constituent at the beginning. I explain that DI functions as an evidential strategy in English, a language which lacks standard evidential markers, and that this specific discursive status explains its structural properties and its main differences from locative inversion. My analysis of DI as an evidential strategy also accounts for its restricted distribution and the otherwise unexpected difference between DI and its non-inverted counterpart in readings of the present/past tense.

In his paper, **Imanol Suárez-Palma** explicitly engages in the debate about the informational status of preverbal subjects, a controversial issue in the current literature regarding the projection of pragmatic information within syntax. He investigates a type of middle-passive sentence in Spanish with a dative possessor in a preverbal position, either as the sole fronted constituent or used in conjunction with the theme DP, which functions as the syntactic subject, i.e., "*A Ismael se le ven las arrugas fácilmente*", "*A Ismael las arrugas se le ven fácilmente*, or "*Las arrugas a Ismael se le ven fácilmente*", with all of these sentences meaning "Ismael's wrinkles are easy to see". Suárez-Palma's analysis posits that the possessor originates within the theme DP and rises to the position of the specifier of an applicative projection in order to be licensed with the dative case. To circumvent the minimality violation that would arise if the preverbal theme DP were to occupy the canonical Spec-TP position, he proposes that both the preverbal lexical dative and the theme DP function as left-dislocated constituents, co-referring with empty pronominals

in argument positions. This proposal then supports the view that preverbal subjects and dative DPs ultimately occupy non-argumental positions in Spanish when they serve as the aboutness-shift topic.

Last but not least, **Julio Villa-García's** paper investigates the syntax of the clausal left edge, parameterizing (and micro-parameterizing) the various realizations of the left-peripheral heads observed in English and Spanish. Specifically, Villa-García explores the lexicalization of the complementizers *that* and *que* in their respective linguistic varieties, demonstrating their presence in nearly all constructions traditionally associated with the CP discourse domain, including exclamative clauses, interrogative contexts, and subjunctive clauses. This study reveals the significant degree of variation in the lexicalization of these complementizers, identifying instances where the complementizer may remain silent (e.g., the high *that* in English), configurations where both the complementizer and the left-peripheral phrase are simultaneously realized (e.g., exclamatives that use *vaya* and *que* in certain varieties of Asturian Spanish), scenarios in which only the left-peripheral constituent is expressed (such as *wh*-interrogatives in Peninsular Spanish), and cases of recomplementation involving an element flanked by overt instances of the complementizer. Through a detailed analysis of these options, Villa-García concludes that the presence or absence of *that* or *que* may be indicative of a complex underlying structure and cannot be merely reduced to a pronunciation parameter. On the contrary, complementizer lexicalization seems to be processing-based, with discourse playing a crucial role in determining these various options across linguistic varieties.

4. Looking Ahead

The issues discussed in this volume represent just a small sample of the various aspects that formal theories must address in order to gain a full understanding of the syntax–pragmatics interface and thus pave the way for future research in this domain.

In studying the syntax–pragmatics interface, linguists aim to clarify the interaction between the computational system and the intentional system. This requires moving away from the strong conception of modularity that was foundational to the Chomskyan generative approach and towards the assumption that the computational system is not blind to discourse. However, to maintain the theory's predictive power and explanatory adequacy, it is essential to avoid an unconstrained syntactic representation of pragmatic information. Thus, much more research is required to establish a sufficient yet restrictive repertoire of pragmatic features and functional categories that should be integrated into the syntactic structure. Additionally, defining all of the constructions in which these pragmatic features play a crucial role remains crucial, specifically constructions in which the constituents are organized based on their relative salience in discourse or their contribution to illocutionary values. Finally, to provide new insights into the interplay between grammatical and pragmatic competence, the range of languages examined must be broadened, and larger samples of data must be sourced.

Acknowledgments: All of the papers in this Special Issue underwent a double-blind peer review process, and I am very grateful to the anonymous reviewers that kindly contributed their time and expertise. I also want to express my deepest gratitude to the Editors of *Language*. Lastly, a posthumous note of gratitude goes to Andrew Radford, from whom many of us have learned so much: this volume is dedicated to him.

Conflicts of Interest: The author declares no conflicts of interest.

References

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

Cruschina, S. (2012). *Discourse-related features and functional projections*. Oxford University Press.

Erteschik-Shir, N. (1997). *The dynamics of focus structure*. Cambridge University Press.

Frascarelli, M., & Hinterhölzl, R. (2007). Types of topics in German and Italian. In K. Schwabe, & S. Winkler (Eds.), *On information structure, meaning and form* (pp. 87–116). John Benjamins.

Haegeman, L. (2014). West Flemish verb-based discourse markers and the articulation of the speech act layer. *Studia Linguistica*, *68*(1), 116–139. [CrossRef]

Hauser, M. D., Chomsky, N., & Fitch, W. T. (2002). The faculty of language: What is it, who has it, and how did it evolve? *Science*, *298*(5598), 1569–1579. [CrossRef] [PubMed]

Krifka, M. (2023). Layers of assertive clauses: Propositions, judgements, commitments, acts. In J. M. Hartmann, & A. Wollstein (Eds.), *Propositionale Argumente im Sprachvergleich: Theorie und Empirie/Propositional arguments in cross-linguistic research: Theoretical and empirical issues* (pp. 116–183). Narr.

Kuroda, S.-Y. (1972). The categorical and the thetic judgment: Evidence from Japanese syntax. *Foundations of Language*, *9*(2), 153–185.

Lobo, M., & Martins, A. M. (2017). Subjects. In A. Dufter, & E. Stark (Eds.), *Manual of Romance morphosyntax and syntax* (pp. 27–88). De Gruyter.

Mendívil-Giró, J. L. (2018). Is universal grammar ready for retirement? A short review of a longstanding misinterpretation. *Journal of Linguistics*, *54*(4), 859–888. [CrossRef]

Miyagawa, S. (2022). *Syntax in the treetops*. MIT Press.

Ojea, A. (2017). Core intentional features in the syntactic computation: Deriving the position of the subject in Spanish. *Lingua*, *195*, 72–91. [CrossRef]

Pinker, S., & Jackendoff, R. (2005). The faculty of language: What's special about it? *Cognition*, *95*(2), 201–236. [CrossRef] [PubMed]

Rizzi, L. (1997). On the fine structure of the left periphery. In A. Belletti, & L. Rizzi (Eds.), *Elements of grammar* (pp. 281–337). Kluwer.

Ross, J. R. (1970). On declarative sentences. In R. A. Jacobs, & P. S. Rosenbaum (Eds.), *Readings in English transformational grammar* (pp. 222–77). Ginn.

Sasse, H.-J. (1987). The thetic/categorical distinction revisited. *Linguistics*, *25*(4), 511–580. [CrossRef]

Speas, P., & Tenny, C. (2003). Configurational properties of point of view roles. In A. M. DiSciullo (Ed.), *Asymmetry of grammar. Volume 1: Syntax and semantics* (pp. 315–344). John Benjamins.

languages

MDPI

Article

Microvariation at the Interfaces: The Subject of Predication of Broad Focus VS Constructions in Turinese and Milanese

Delia Bentley [1,*] and Francesco Maria Ciconte [2,*]

[1] Department of Linguistics and English Language, The University of Manchester, Manchester M13 9PL, UK
[2] Department of Humanities, Arts and Social Sciences, University of Chieti-Pescara, 66100 Chieti, Italy
* Correspondence: delia.bentley@manchester.ac.uk (D.B.); francesco.ciconte@unich.it (F.M.C.)

Abstract: Presentational constructions, i.e., structures which introduce an event into the universe of discourse, raise the question of what it means for a predication to be entirely new in information structural terms. While there is growing consensus that these constructions are not topicless, there is no agreement on how to analyse their topic. The Romance languages of Northern Italy have figured prominently in this debate because the presentational constructions of many such languages exhibit VS order and an etymologically locative clitic in subject clitic position. This clitic has been claimed to be a subject of predication in a syntactic subject position. Adducing primary comparative evidence from Milanese and Turinese, we discuss patterns of microvariation which suggest that the etymologically locative clitic need not be a syntactic subject and can mark an aboutness topic provided by the discourse situation alone. We propose a parallel-architecture, Role and Reference Grammar account whereby the microvariation under scrutiny is captured in terms of the interfaces that are involved in the parsing of utterances. This account considers discourse to be an independent module of grammar, which, alongside the semantic and syntactic modules, is directly involved in linguistic variation and change.

Keywords: aboutness topic; interfaces; microvariation; parallel architecture; presentational construction; Role and Reference Grammar; Romance; subject clitic

Citation: Bentley, Delia, and Francesco Maria Ciconte. 2024. Microvariation at the Interfaces: The Subject of Predication of Broad Focus VS Constructions in Turinese and Milanese. *Languages* 9: 37. https://doi.org/10.3390/languages9020037

Academic Editor: Ana Ojea

Received: 1 November 2023
Revised: 12 January 2024
Accepted: 16 January 2024
Published: 24 January 2024

1. Introduction

Many dialects of Northern Italy exhibit an etymologically locative clitic in sentences with VS order which introduce a new event into the universe of discourse (Tortora 1997, 2014; Parry 2000, 2013; Manzini and Savoia 2005, vol. 2; Ciconte 2008, 2011; Pescarini 2016, p. 749; Bentley 2018; Flecchia 2021, 2022; Bentley and Cennamo 2022).[1] We call these sentences *presentational constructions* (Parry 2013, p. 511), and we refer to the etymologically locative clitic as *presentational clitic*. Milanese *ghe* is a relevant example.[2]

(1) **Gh'** è rivà i to surèi. (Milanese)
PRESCL be.3SG arrive.PSTP the your sisters
'Your sisters have arrived.'

The presentational clitic occurs not only with verbs of motion, like *rivàr* 'arrive' in (1), but also with verbs that lack a locative argument, like *murìr* 'die' in (2).

(2) **Gh'** è mort tanti suldà. (Milanese)
PRESCL be.3SG die.PSTP many soldiers
'There died many soldiers.'

This suggests that rather than encoding a location, *ghe* signals a property of presentational focus, which deserves investigation. Comparing the development of the presentational clitic to that of the existential proform 'there' (Ciconte 2008, 2011), the existing analyses argue that the presentational clitic, originally a resumptive locative pronoun, developed into a subject agreement marker, i.e., the marker of verbal agreement with a

covert locative subject of predication (see Parry 2013 for the diachronic account; for the clitic in synchronic terms, see Burzio 1986; Saccon 1992; Parry 1997, p. 243; 2000; Tortora 1997; 2014, pp. 29–32). The subject of predication is understood as a spatio–temporal location, which is required by the presentational construction (Benincà 1988; Calabrese 1992; Saccon 1992). The analyses of the clitic as a subject agreement marker are supported by the observation that the presentational clitic takes the position of a subject clitic, i.e., a bound pronominal form, which is an extended exponent of subject–verb agreement. In fact, the presentational clitic clusters with or occurs in complementary distribution with subject clitics.[3] The verb also fails to agree in number with the postverbal NP, which suggests that the latter is not the subject of the presentational construction: see the number mismatch between the singular auxiliary *è* 'is' and the plural NPs *i to surèi* 'your sisters' and *tanti suldà* 'many soldiers' in (1) and (2), respectively.

Comparative evidence from other dialects of Northern Italy reveals, however, that there is variation in the position of the presentational clitic and in the agreement relation between the verb and the postverbal NP. Thus, Turinese *je* contrasts with Milanese *ghe*, in that it does not figure in a subject–clitic position (cf. 3 and 4) and it is not ruled out when the verb agrees in number with the postverbal NP (cf. 4).[4]

(3) A l' è riva**je** toe sorele. (Turinese)
SCL.3SG AUXCL be.3SG arrive.PSPT.PRESCL your sisters
'Your sisters have arrived (here/where I am/was).'

(4) A **son** rivà**je** toe sorele. (Turinese)
SCL.3PL be.3PL arrive.PSTP.PRESCL your sisters
'Your sisters have arrived (here/where I am/was).'

The evidence in (3) and (4) challenges the view that the presentational clitic satisfies a syntactic subjecthood requirement. However, it does not conflict with an analysis of the clitic as the expression of the deixis of the discourse situation in which the new event is announced and to which it is relevant. In such cases, the discourse situation is the aboutness topic of the utterance, or what the utterance is about (Gundel 1988), and the presentation clitic indexes it by virtue of its deictic features. Supported by the speaker-oriented deixis of presentational constructions with *je* (see the translation of examples 3 and 4), this hypothesis draws on the notion that, despite introducing a new event, apparently topicless sentences require a stage topic (Erteschik-Shir 1997, p. 8) or, otherwise put, a subject of predication (Saccon 1992; Benincà 1988; Calabrese 1992; Bianchi 1993; Parry 2013).

In this article, we examine first-hand Turinese and Milanese evidence collected in loco with questionnaire-assisted interviews. We claim that the microvariation in the presentational constructions of the two dialects resides at the interfaces that are involved in sentence parsing. We frame our account in terms of Van Valin's (2023, pp. 123–25) syntax–semantics linking algorithm, which is an idealization of the hearer's perspective in linguistic communication. The presentational clitic can be understood as a locative argument of a verb of motion, in which case it is linked from syntax to a position in the semantic representation of the clause. Alternatively, the clitic indexes the discourse situation and is directly linked from syntax to discourse. This latter interpretation is not restricted to presentational constructions with verbs of motion. To capture the variation attested in Turinese (cf. 3 and 4), we advance the hypothesis that a new process of grammaticalization is under way, where referential, i.e., locative, *je*, which was never ousted from the system, is being reanalysed for a second time. As a result of the reanalysis, *je* becomes an index of the aboutness topic of the utterance, which is the discourse situation.

Both the construal of *je* as a locative argument and that as an index of the discourse situation are compatible with the establishment of an agreement relation between the verb and the postverbal NP, the notion of subjecthood being broken down, in our analysis, into an aboutness relation, which is not in principle syntactic and can be satisfied in discourse, and an agreement relation, which is syntactic and construction-specific (Van Valin and LaPolla 1997, pp. 242–309; LaPolla 2023). The construction-specific agreement relation is missing obligatorily in the Milanese presentational construction with the presentational

clitic (cf. 1 and 2), and optionally in its Turinese counterpart (cf. 3 and 4). In the absence of V–S agreement, the grammars of both dialects require that the aboutness relation be expressed overtly by the presentational clitic (cf. 1–3). We take this to be a constructional requirement of the autochthonous presentational pattern of these dialects. Instead, the presentational construction with V–S agreement is a result of contact with and pressure from Italian.

In the sections to follow, we discuss the micro-typology of presentational constructions found in Milanese and Turinese (Section 2). We then briefly introduce Parry's (2013) analysis of the diachrony of presentational clitics (Section 3.1), and we advance a hypothesis on the variation observed in Turinese nowadays (Section 3.2). We introduce our framework in Section 4, and we propose our formal analysis in Section 5. Brief conclusions are drawn in Section 6.

2. Microvariation in the Presentational Construction

The Milanese and Turinese clitics introduced in the previous section illustrate two of the three formal types of presentational clitic attested in the dialects of Northern Italy: proclitic (cf. 1 and 2), enclitic (cf. 3 and 4), and sequential (cf. 5). This micro-typology is sketched in Table 1.

(5)

Ngh'	è	gnö	denti-**ghi**	na	segretaria	(Borgomanero)
LOC	be.3SG	come.PSTP	inside-LOC	a	secretary	
int	la	stônza				
in	the	room				

'A secretary entered the room.'
(Tortora 2014, p. 20)

Table 1. A micro-typology of presentational clitics.

Type I: Proclitic	Type II: Enclitic	Type III: Sequential
Milanese *ghe* (cf. 1 and 2)	Turinese *je* (cf. 3 and 4)	Borgomanerese *ngh'... gghi* (cf. 5)

An etymologically locative clitic is known to have undergone partial univerbation in some of the Northern dialects, figuring on the inflected forms of 'have' and 'be' under conditions outlined in Benincà (2007).

(6)

Gh'ò	cantà.		(Venetan)
CL-have.1SG	sing.PSTP		

'I have sung.'
(Benincà 2007, p. 28)

Given that our focus is on presentational constructions, we shall leave this last form out of the discussion. In what follows, we introduce our findings on Milanese and Turinese, which exhibit clitics of types I and II.

2.1. Our Findings

We conducted questionnaire-assisted interviews with Milanese and Turinese speakers and found that, in both dialects, the absence of agreement is normally dependent on two conditions: (i) S must not be a personal pronoun, and (ii) V must be a Vendlerian state, achievement, or accomplishment.[5] Thus, the second person plural pronoun obligatorily controls person and number agreement on the verb in (7) and (8).[6]

(7)

Dopu	si	/	*gh	/	*l'	*è	rivà	vuialter.	(Milanese)
after	be.2PL		PRESCL		AUXCL	3SG	arrive.PSTP	you.PL	

'Then you arrived.'

(8) Peui i seve vnü / *a *l'è (Turinese)
after SCL.2PL be.2PL come.PSTP SCL.3SG AUXCL-be.3SG
vnü(je) voi.
come.PSTP(PRESCL) you.PL
'Then you came.'

In (9) and (10), V–S agreement is, instead, required because the verb *balé* 'dance' is a Vendlerian activity (see Parry 2013; Bentley 2018; Bentley and Cennamo 2022 for further detail).

(9) (A la festa) an balà / *l' / *gh' a (Milanese)
at the party have.3PL dance.PSTP AUXCL PRESCL have.3SG
balà i to gent.
dance.PSTP the your parents
'(At the party) your parents danced.'

(10) (A la festa) a l'an balà / *a (Turinese)
at the party SCL.3PL AUXCL-have.3PL dance.PSTP SCL.3SG
l'a balà(*je) tò papà e toa mama.
AUXCL-have.3SG dance.PSTP(PRESCL) your dad and your mum
'(At the party) your mum and dad danced.'

The second condition appears to be sporadically violated in Turinese, a point to which we return in Section 3.2.

Apart from these shared constraints, in Milanese we found two patterns. The one illustrated in (11) was prevalent: the auxiliary hosts the presentational clitic and does not agree in number with the postverbal noun phrase.

(11) a. *(**Gh**')è rivà i to surèi / di pac. (Milanese)
PRESCL-be.3SG arrive.PSTP the your sisters of parcels
'Your sisters have arrived'/'There arrived some parcels.'

b. *(**Gh**')è mort tanti suldà.
PRESCL-be.3SG die.PSTP many soldiers
'There died many soldiers.'

V–S agreement was also attested, but the presentational clitic turned out to be incompatible with the agreeing verb.

(12) a. (*Gh')in rivà i to surèi / di pac. (Milanese)
PRESCL-be.3PL arrive.PSTP the your sisters of parcels
'Your sisters have arrived.'/'There arrived some parcels.'

b (*Gh')in mort tant sulda.
PRESCL-be.PL die.PSTP manysoldiers
'There died many soldiers.'

Subject clitics are known to be incompatible with preverbal object clitics in some Friulian and Francoprovençal varieties (Benincà and Vanelli 1986; Roberts 1993; Poletto and Tortora 2016, p. 785). We found that the presentational clitic could not occur in the presence of a reflexive clitic, and, in this case, the verb obligatorily agreed in number with the postverbal NP.

(13) Varda: s'in s-cepà tanti ram. (Milanese)
look.IMP.2SG REFL-be.3PL break.PSTP many tree-branches
'Look! Many branches have broken.'

In sum, the Milanese presentational clitic *ghe* is required in VS presentationals without verb agreement and is incompatible with such agreement. Taking V–S agreement and the presence of the presentational clitic to be two binary variables, which we call {±Agr} and {±Cl}, the situation found in Milanese presentational constructions can be represented as in Table 2.

Table 2. Milanese presentational constructions: two patterns.

	Pattern (i)	**Pattern (ii)**
{Agr}	–	+
{Cl}	+	–

At this juncture, a brief digression is necessary. We should note that Milanese *ghe* does co-occur with the inflected copula of locative and existential constructions exhibiting a postcopular personal pronoun, as we illustrate here.

(14) Chi l'è che gh'è in cüsina? (Milanese)
who AUXCL-be.3SG that LCL-be.3SG in kitchen
Ghe sun mi, in cüsina.
LCL be.1SG I in kitchen
'Who is the kitchen? I am the kitchen (lit., There am I, in the kitchen).'

(15) Maria l'è no in de per lé: (Milanese)
Mary SCL-be.3SG NEG in by for her
ghe sun mi.
PF be.1SG I
'Mary is not alone: I am there for her (lit., There am I).'

The *there* form of locatives and existentials must, however, be distinguished from the presentational clitic on both empirical and theoretical grounds. First, both in Italian and in some northern dialects of Italy not discussed in depth here, an etymologically locative clitic occurs in locatives and existentials (cf. 16 and 17) but is unattested in presentationals (cf. 18).

(16) a. (In cusina) **gh** son mi. (Grosio)
in kitchen LCL be.1SG I
'I am in the kitchen (lit., In the kitchen there am I).'

b. (In cucina) **ci** sono io. (Italian)
in kitchen LCL be.3PL I
'I am in the kitchen (lit., In the kitchen there am I).'

(17) a. Maria l'é miga de per lé: (Grosio)
Mary SCL-be.3SG NEG by for her
ghe son mi.
PF be.1SG I
'Mary is not alone: I am there for her (lit., There am I).'

b. Maria non è sola: **ci** sono io. (Italian)
Mary NEG be.3SG alone PF be.1SG I
'Mary is not alone: I am there for her (lit., There am I).'

(18) a. L'é rivä i toa sureli. (Grosio)
SCL-be.3SG arrive.PSTP the your sisters
'Your sisters have arrived.'

b. Sono arrivate le tue sorelle. (Italian)
be.3PL arrive.PSTP.PL the your sisters
'Your sisters have arrived.'

While the proform in (16) resumes the previously introduced locative predicate *in cusina* 'in the kitchen', hence the gloss LCL 'locative clitic', the existential proform in (17), glossed PF, has been argued to signal the context dependence of existential sentences, which are predications of an implicit contextual argument (Bentley et al. 2015, p. 146; following Francez 2007). In contrast, the analysis of the presentational clitic is not similarly straightforward. The presentational construction is fully interpretable in its absence and is truth-conditionally equivalent with the corresponding SV sentence (Lambrecht 1988, p. 115;

Karssenberg 2016; 2018a, p. 23; 2018b). Furthermore, depending on the semantics of the verb, the presentational clitic can receive a locative interpretation (Section 1). In the present work, we shall, therefore, leave aside existentials and locatives and focus on presentational constructions, as defined in Section 1, because it is in such constructions that ambiguity arises in the interpretation of the etymologically locative clitic, revealing the key role of the interfaces in microvariation.

Turning now to Turinese, to begin with, not one but two etymologically locative forms are found in the presentational constructions of this dialect (Parry 1997, 2000, 2010, 2013). At first, the two forms appear to be allomorphs of the same morpheme (see, e.g., Parry 2013, p. 514): the one, *j*, occurs proclitically to the finite verb in the simple tenses (cf. 19), whereas the other, *je*, occurs enclitically to the participle (cf. 20) or the infinitive.

(19) Se a-**j** seurt ël sol, sì a-**j**
if EXPL-PRESCL come.out.3SG the sun here EXPL-PRESCL
nass ji bolè.[7]
be.born.3SG the mushrooms
'If the sun comes out, mushrooms will appear here.'
(Parry 2013, p. 515, data from Burzio 1986)

(20) Che bel! A l'è na(ssù)**je** (Turinese)
what beautiful SCL.3SG AUXCL-be.3SG be.born.PSTP.PRESCL
le fior.
the flowers
'How nice! The flowers have appeared.'

On further inspection, this hypothesis would appear to be problematic. Unlike *je*, *j* is not constrained in terms of the Aktionsart of the verb and, when occurring with activity verbs, it is associated with habitual or iterative aspect (Parry 2013, p. 541).

(21) An cost let a-**j** deurm mie fije. (Turinese)
in this bed EXPL-PRESCL sleep.3SG my daughters
'This bed is where my daughters sleep.'
(Parry 2013, p. 541)

In addition, the pattern with *j* was not normally chosen by our informants. When it was chosen, *j* was hosted by an inflected form of the verb (for the latter point, see also Tosco et al. 2023, p. 184).

(22) Guardoma la partita e a(-**j**) intro (Turinese)
watch.1PL the game and SCL.3PL(-PRESCL) enter.3PL
doi lader dal giardin.
two thieves from-the garden
'We are watching the game and two thieves enter from the garden.'

The evidence in (22) contrasts with that reported in the literature (cf. 19, 21), including the treatments of earlier stages of the language (Parry 2013, p. 539), where *j* correlates with lack of V–S agreement.

In contrast with *j*, *je* is very well attested in our dataset, and its distribution vis-à-vis V–S agreement is not the same as that found with *j*. In fact, we found the threefold possibility illustrated in Table 3. The fourth logical combination of the two variables, the lack of both V–S agreement and *je*, was not attested in Turinese.

Table 3. Turinese presentational constructions with *je*: three patterns.

	Pattern (i)	**Pattern (ii)**	**Pattern (iii)**
{Agr}	−	+	+
{Cl}	+	−	+

Patterns (i) and (ii) of Table 3 are the same as those found with *ghe* (see Table 2), whereas according to our records, pattern (iii) is unknown to Milanese. The three combinations of the two binary variables are illustrated in the following examples from Turinese.

[pattern (i)]

(23) a. A l'è riva**je** toe sorele
SCL.3SG AUXCL-be.3SG arrive.PSTP.PRESCL your sisters
di pachet.
of parcels
'Your sisters have arrived'/'There arrived some parcels.'

b. A l'è nassù**je** tante fior.
SCL.3SG AUXCL-be.3SG be.born.PSTP.PRESCL many flowers
'Many flowers have appeared.'

[pattern (ii)]

(24) a. A son montà ën païs i tòi
SCL.3PL be.3PL go.up.PSTP in village the your
nòno.
grandparents
'Your grandparents have gone/come up to the village.'

b. A son calà i sgnor dël
SCL.3PL be.3PL come.down.PSTP the people of.the
pian ëdzora.
floor of.upstairs
'The people from the upstairs floor have come down.'

[pattern (iii)]

(25) a. A **son** rivà**je** toe sorele /
SCL.3PL be.3PL arrive.PSTP.PRESCL your sisters
di pachet.
of parcels
'Your sisters have arrived.'/'There arrived some parcels (here, where I am).'

b. A **son** nassu**je** tante fior.
SCL.3PL be.3PL be.born.PSTP.PRESCL many flowers
'Many flowers have appeared.'

Three observations on the patterns in (23)–(25) are in order. Firstly, according to the literature, pattern (i) is the autochthonous Turinese presentational construction (Parry 1997, p. 243; 2013, pp. 514–15; Flecchia 2022, p. 44). Pattern (ii), instead, is the outcome of convergence between the grammar of Turinese and that of the more prestigious language Italian (Flecchia 2022), which has no presentational clitic and requires V–S agreement in presentationals (cf. 18b).

Secondly, *je* occurs in complementary distribution with any non-subject clitics that the sentence may require. In (26), we provide an example with the direct object clitic *lo* 'it'; the position of *je* is the same as that of this clitic. In (27) we report a presentational construction, where the participle hosts a dative and a partitive clitic. *Je* is banned with consequent obligatoriness of V–S number agreement.

(26) Col film, a l'an vist-lo tuti (Turinese)
that film SCL.3PL AUXCL-have.3PL see.PSTP-OCL all
i mè amis.
the my friends
'That film, all my friends have seen it.'

(27) A son rivà-(*je)-m-ne(*-je) doi. (Turinese)
SCL.3PL be.3PL arrive.PSTP(*PRESCL).DATCL.PARTCL(*PRESCL) two
'There arrived two of them to me.'

Thirdly, according to many informants *je* has a speaker-oriented deictic function, indicating that the location of the event is the same as that of the speaker, and suggesting first-hand witnessing of the event.

In sum, abstracting away from the shared constraints cited at the beginning of this section (cf. 7–10), a more varied array of patterns was found in the presentational constructions of Turinese than in those of Milanese. In the latter dialect, the presentational proclitic *ghe* is hosted by a verb that will exhibit invariant third-person singular morphology, regardless of the number feature of the postverbal NP. Although V–S number agreement is also an option, it is mutually exclusive with the presentational clitic. As for Turinese, in contrast with what is reported in the literature, we found the proclitic *j* to occur rarely in presentational constructions and to have no effect on the number agreement between the verb and the postverbal NP, which is regularly found in the present. The other form known to occur in Turinese, *je*, attaches enclitically to the participle of the perfect, on a par with non-subject clitics. While the literature reports that *je*, like Milanese *ghe*, is incompatible with V–S agreement, we found this not to be the case: *je* and the number agreement specifications on the perfect auxiliary are not mutually exclusive.

3. The Development of the Presentational Clitic and Its Theoretical Consequences

3.1. The Diachrony of J*: Parry's Account*

The development of the presentational clitic is well documented and has been studied in some depth. An etymologically locative form is attested in early Italo–Romance existentials since the 12th century, its reanalysis into a non-referential existential proform dating from the 13th–14th centuries (Ciconte 2008, 2011, 2013). To capture this development, Parry (2013) has claimed that the existential proform, which was not found in Latin existentials, originated as a locative pronoun which resumed an extra-clausal locative phrase or a locative phrase that was interpolated between the verb and its postverbal argument: see the [V-Loc-NP] order in (28).

(13th c. Veronese, Giacomino da Verona, *Babilonia*)

(28)	Asai	**g'**è	**là**	**çó**	bisse [. . .]
	many	there-is	there	down	grass-snakes

'There are many grass-snakes [. . .] down there.'
(Parry 2013, p. 530)

In the V2 syntax of old Romance, which was characterised by (X)VS order, the postverbal position was the default position of the subject. Therefore, in the structures where the locative phrase was interpolated between the verb and its postverbal argument, the locative phrase lent itself to being analysed as the subject. In turn, the co-referring clitic could be reanalysed as a subject agreement marker. According to Parry, the existential proform originated from this reanalysis.

In Piedmontese, the presentational clitic J (*j*/*je*) did not spread to presentational constructions until the 17th century, and even then, it was only found with the verb *arivé* 'arrive' (Parry 2013, p. 539). By the 18th century, it was attested with a larger range of verbs of directed motion and change of state, as can be seen below.

(18th c. Turinese, I. Isler, ed. Viglongo 1968)

(29)	L'	é	bin	dal	liam	ch'	a**i**	nass
	it	is	well	from.the	manure	that	EXPL.SCL-LOC.CL	be.born.3SG
	le	fior.[8]						
	the	flowers						

'Indeed, it's from manure that flowers grow.'
(Parry 2013, p. 539)

According to Parry, in this structure a referential locative clitic was also reanalysed as a subject agreement marker: "Indeed, it was a similar process of syntactic reanalysis that produced subject–clitic agreement markers. The latter were originally used as clause–internal resumptive pronouns linked to dislocated subjects, but later weakened into compulsory

agreement markers on the verb (see for example, Poletto 1993, and for Piedmontese, Parry 1993)." (Parry 2013, p. 529).

This reanalysis was favoured by the postverbal occurrence of the argument (see *le fior* 'the flowers' in 29), and the failure of the verb to enter into an agreement relation with it (*nass* is singular in 29). Parry's claim is, therefore, that the locative clitic was reanalysed as an agreement marker, on a par with subject clitics. The presentational clitic thus came to be associated with a subject position, specifically, the position which Cardinaletti (2004) calls SubjP (Parry 2013, pp. 535–36), retaining some locative value at the level of discourse and spelling out the locative subject of predication of the presentational construction.[9]

3.2. The Re-Grammaticalization of je

At first, the evidence uncovered in our recent survey of Turinese appears to jar with the notion that a locative pronoun developed into a subject marker J, with allomorphs *j*/*je*. The challenge for this hypothesis is twofold. On the one hand, we found that both *j* and *je* occur in VS constructions where the verb patently agrees in number with the postverbal NP (cf. 22, 25a,b). This fact is irreconcilable with the assumption that *j* and *je* signal that the verb agrees with a subject of predication that is different from the postverbal NP. On the other hand, *j* hardly occurs in our dataset, whereas *je* figures not only in the autochthonous pattern known from the literature (cf. 23a,b), but also in a presentational pattern with V–S agreement, which seems to have gone unnoticed so far (cf. 25a,b).

Capitalising on the observation that patterns (ii) and (iii) from Table 3 (cf. 24, 25a,b) were not traditionally found in Turinese, we propose that the new evidence gathered in our survey does not challenge Parry's (2013) hypothesis on the development of a locative pronoun into a presentational clitic, but rather suggests that new variation is available in contemporary Turinese grammar, and a new development may be under way. Starting from the outcome of the diachronic development discussed in Parry (2013), i.e., pattern (i) from Table 3 (cf. 30a), a new presentational pattern was introduced into the system because of contact with Italian (Ricca 2008; Flecchia 2022), namely pattern (ii) (cf. 30b). Our hypothesis is that J had never lost its locative meaning, but rather had undergone layering, maintaining its old function at the same time as the presentational one (for layering, see Hopper and Traugott 1993, pp. 36, 124–26; for locative j, see Tosco et al. 2023, p. 184). Locative J was introduced into pattern (ii) as the spell-out of a locative argument or to resume a locative adjunct. This is how pattern (iii) originated.

(30)
a. **Stage 1**: Pattern i {+J; −V–S Agr}.
b. **Stage 2**: Pattern i {+J; −V–S Agr}; pattern ii {−J; +V–S Agr}.
c. **Stage 3**: Pattern i {+J; −V–S Agr}; pattern ii {−J; +V–S Agr}; pattern iii {+J; +V–S Agr}.

That J can be locative is suggested not only by the speakers' observation that *je* can have a deictic flavour, signalling the location of the speaker, but also by the few exceptions to the Aktionsart constraint discussed in Section 2.1 (cf. 10). According to our informants, such violations of the ban on activity predicates in presentational focus are only allowed if the location of the event is specified, as can be seen in (31). This structure was provided spontaneously by several informants.

(31)	**A**	**la**	**scola**,	a	l'an	durmì**je**	(Turinese)
	at	the	school	SCL.3PL	AUXCL-have.3PL	sleep.PSTP-LCL	
	tanti		cit.				
	many		children				

'At school, many children have slept (there) (lit. there slept many children).'

The locative interpretation also arises naturally from examples like (25a), repeated in (32) for convenience, where the verb describes directed motion.

(32) **A** **son** rivà**je** toe sorele / (Turinese)
SCL.3PL be.3PL arrive.PSTP.PRESCL your sisters
di pachet.
of parcels
'Your sisters have arrived.'/'There arrived some parcels (here, where I am).'

However, pattern (iii) is also attested, if more rarely, with verbs that do not take a locative argument, like *nasse* 'be born' and *dimagrì* 'lose weight'.

(33) a. A son nassu**je** tante fior. (Turinese)
SCL.3PL be.3PL be.born.PSTP-PRESCL many flowers
'Many flowers have appeared.'

b. A son dimagrì**je** tanti cit.
SCL.3PL be.3PL lose.weight.PSTP.PRESCL many children
'Many children have lost weight.'

This suggests that the locative pronoun of pattern (iii) is subject to reanalysis in synchrony and, as a consequence, is being re-grammaticalized into the marker of a more abstract property of the construction, which is only loosely related to locative meaning. To understand what this property might be, it is important to return to the point made previously that *j*, the allomorph of J expected to figure in the simple tenses, makes only a few sporadic appearances in our dataset. In other words, in the examples in the present tense, pattern (ii) was normally chosen. The optionality of *j* was shown in (22); in (34), we report the present-tense pattern that was by far predominant in the responses to our questionnaire.

(34) a. A-j è la guera: (Turinese)
SCL.3SG-PF be.3SG the war
a **moero** tanti soldà.
SCL.3PL die.3PL many soldiers
'There's a war: many soldiers are dying.'

b St'ane sì **a** **naso** poche masnà.
this-year here SCL.3PL be.born.3PL few children
'Only few children have been born this year.'

From the contrast between the present and the perfect, we conclude that the erstwhile allomorphs of J have parted ways, with *j* becoming virtually obsolete, and *je* acquiring a new function of its own, a function which emerges in a construction that requires the perfect, where *je* was always licensed, namely the presentational construction. More specifically, *je* is reanalysed as the marker of the deixis of the discourse situation in which the announcement of a new event is made.[10] Since it is compatible with the agreement specifications on the verb, it must be concluded that it is not a syntactic subject. However, in light of the obligatoriness of *je* in the absence of agreement, it must be the case that the presentational clitic is a constructional requirement in Turinese (as in Milanese), when the construction is unambiguously presentational.[11]

Since we assume that the last stage in the development outlined in (30) involves a new role for *je*, which has parted from *j*, we represent the variation attested in contemporary Turinese as follows.

(35)
Pattern (i) {+*je*; −V–S Agr} ~ pattern (ii) {−*je*; +V–S Agr}~ pattern (iii) {+*je*; +V–S Agr}

After introducing the framework adopted in our analysis (Section 4), in Section 5 we shall provide a formal account of this variation, outlining the synchronic conditions for the reanalysis of *je*.

4. Role and Reference Grammar

Role and Reference Grammar (henceforth RRG; see Foley and Van Valin 1984; Van Valin and LaPolla 1997; Van Valin 2005, 2023; Bentley et al. 2023) is a *parallel architecture* theory (Jackendoff 2002, pp. 125–30), which represents discourse–pragmatics and semantics separately from syntax and seeks explanation in the interplay of these independent modules of grammar. A bidirectional algorithm governs the mapping—or *linking*—of semantics with syntax, in language production, and syntax with semantics, in language comprehension. Regardless of direction, the linking consists of two phases: the lexical phase builds the meaning of the clause, starting from the lexical meaning of the predicators, and assigns macroroles, i.e., generalized semantic roles, to the arguments, following universal principles that are grounded in Dowtyan lexical decomposition rules (Van Valin and LaPolla 1997, pp. 139–78; Van Valin 2023, p. 113). The morphosyntactic phase determines the morphosyntax of the clause and is characterized by a great deal of cross-linguistic variation to do with the assignment of grammatical relations, voice alternations, alignment, head- vs. dependent-marking orientation, etc. (Van Valin and LaPolla 1997, pp. 242–309; Van Valin 2023, pp. 147–49). The linking is paralleled by the discourse–pragmatic dimension, which can intervene at any point, although its role is most acutely felt in the morphosyntactic phase, resulting in significant intra- and cross-linguistic variation (Bentley 2023b). The relation between the three dimensions is represented schematically in Figure 1.

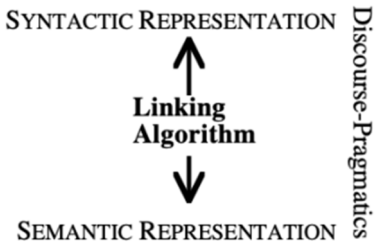


Figure 1. The interplay of syntax, semantics and discourse pragmatics (Van Valin 2023, p. 19).

Grammatical relations are not universals of syntactic theory for RRG; nor is there a subjecthood well-formedness requirement on the clause (LaPolla 2023). In the languages which do have grammatical relations, these are defined in terms of construction-specific restricted neutralisations of semantic roles for syntactic purposes. Thus, taking number verbal agreement to be a construction, this is a grammatical relation in the languages under discussion because it neutralises the distinction between A (actor of transitive), S (actor and undergoer of intransitive) and d-S (derived S in the passive), leaving out U (undergoer of transitive). Actor and undergoer are macroroles, or generalised semantic relations, in RRG, a point to which we return below. What matters here is that the contrast between the two is neutralized by number agreement but leaves out U, undergoer of transitive. The grammatical relation which gathers {A, S, d-S}, leaving out U, is called a P(rivileged) S(yntactic) A(rgument).

Albeit formed in accordance with the general principles of clause construction, tree structures are templatic and stored in language-specific inventories. A parser intervenes early in the syntax–semantics linking to output a tree structure for the input received (Van Valin 2023, pp. 116–25). Neither movement nor empty positions are allowed, and therefore, tree structures must represent the actual order of the elements in the clause, thus satisfying a principle which is normally referred to as the concreteness constraint.

An important property of the RRG theory of grammar is that it is both projectionist and constructional (Bentley 2023a). It is projectionist in the sense that it derives key aspects of the syntax of the clause from facets of lexical meaning. It is constructional insofar as it assumes that competence in a language includes the knowledge of its constructions. Therefore, alongside the Syntactic Inventory and the Lexicon, the grammar of that language will include an inventory of Constructional Schemas, which are constellations of instructions for the formation and the parsing of each of the constructions of that language. In due course,

this aspect of the RRG conception of grammar will become relevant to our discussion. The general organization of grammar discussed thus far is illustrated in Figure 2.

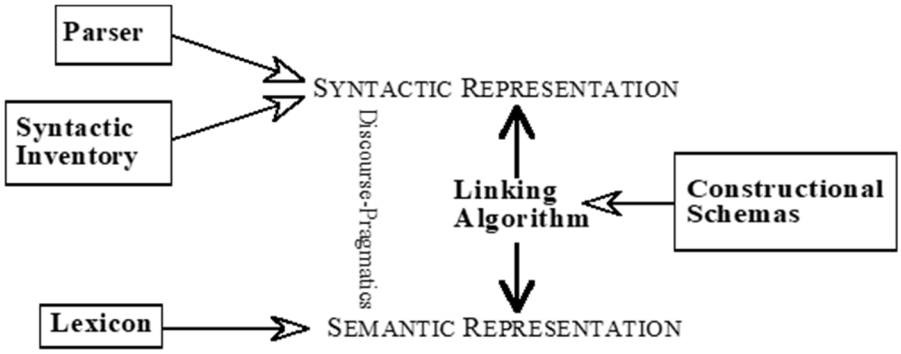


Figure 2. Organization of grammar in RRG (Van Valin 2005, p. 134).

A key principle governing the linking algorithm is the Completeness Constraint, which states that "[a]ll of the arguments explicitly specified in the semantic representation of a sentence must be realized syntactically in the sentence, and all of the referring expressions in the syntactic representation of a sentence must be linked to an argument position in [. . .] the semantic representation of the sentence." (Van Valin 2023, p. 116). While ensuring that every well-formed sentence is interpretable, this principle poses a challenge in cases of null anaphora (prop drop, object drop, silent predicates, etc.), since, as we said, the RRG representation of syntactic trees is constrained by a ban on phonologically null elements. Given that the modules of grammar can link directly with each other, RRG analyses null anaphora in terms of the direct linking of an argument or a predicate in the semantic representation of the clause with its representation in discourse without a concomitant link to syntax (Van Valin 2005; Shimojo 2008).

The RRG representation of discourse draws on von Heusinger's (1999) Discourse Representation Structures, which in turn build on Kamp and Reyle's (1993) Discourse Representation Theory. Discourse Representation Structures aim to capture the incrementality of information in discourse. They include variable-value pairs for the referring expressions that are gradually introduced into discourse and a representation of the semantic propositions in which the variables figure. As new propositions are gradually introduced into discourse, the co-reference relations between the variables of these propositions and those that were introduced previously are also represented. We shall provide examples of Discourse Representation Structures in Sections 5.1 and 5.2, although the issue of co-reference will not be relevant for our purposes, given than we deal with all-new utterances.

The direct linking of semantic and discourse representation that is assumed in RRG to capture null anaphora satisfies the *Extended* Completeness Constraint (Van Valin and Latrouite 2023, p. 496), which ensures that the sentence is fully interpretable without there being any null positions in syntax. The analysis which we shall develop in the next section will demonstrate another application of the Extended Completeness Constraint, whereby the direct linking connects discourse with syntax without involving the semantic representation built from the predicate(s) in the Lexicon.

5. Microvariation at the Interfaces

How can the synchronic variation discussed in previous sections be captured, and how is the change hypothesised for Turinese triggered and enabled? For ease of exposition, we illustrate the relevant variation again below.

(36)
Pattern (i) {+*je*; −V–S Agr} ~ pattern (ii) {−*je*; +V–S Agr} ~ pattern (iii) {+*je*; +V–S Agr}

As the reader will recall, patterns (i) and (ii) were found both in Milanese and in Turinese, although in the former dialect pattern (i) was predominant. The key difference

between the two dialects emerged instead from pattern (iii), which did not occur in Milanese but was well attested in Turinese. It is in this pattern, where the clitic co-occurs with V–S agreement, that the conditions for ambiguity in interpretation arise and it is this ambiguity that enables the new reanalysis of the clitic from locative to presentational (i.e., from a referential locative pronoun to the marker of the deixis of the discourse situation and hence an index of the aboutness topic of the utterance). The trigger of the reanalysis is the presentational construction itself because this construction does not provide an argument which can serve as an aboutness topic for the predication, but a topic is required by all propositions (Erteschik-Shir 1997). In what follows, we shall therefore analyse the two possible interpretations of the clitic and of pattern (iii), taking the perspective of the syntax–semantics linking, which is an idealization of the hearer's viewpoint in communication (Section 5.1). We shall then compare pattern (iii) with patterns (i) and (ii) (Section 5.2) and make some theoretical observations arising from this comparison (Section 5.3).

5.1. Pattern (iii) at the Syntax-Semantics Interface

Consider (37), which is a simplified version of (25a), and assume that this is an utterance that occurs out of the blue: there is no presupposition that x has arrived or that your sisters have done y.

(37)	**A**	**son**	rivà**je**	toe	sorele.	(Turinese)
	SCL.3PL	be.3PL	arrive.PSTP.PRESCL	your	sisters	
	'Your sisters have arrived (here/where I am).'					

In accordance with the algorithm which governs the syntax–semantics interface (Van Valin 2023, pp. 123–25), once the input in (25) is received, the parser outputs a labelled tree structure (Step 1 in Figure 3). This will consist of a Nucleus, hosting the verb and the agreement specifications in the AG(reement) (Inde)X node, and R(eference) P(hrase)s for the pronoun *je* and the NP *toe sorele*. At this point, as much information is gleaned as is possible from the morphosyntax of the clause (Step 2). Given that the verbal inflection in the AGX node is in the third person plural, the clitic *a* is interpreted as a third person plural subject clitic (see note 4), and the plural NP *toe sorele* 'your sisters' is analysed as the controller of agreement or Privileged Syntactic Argument (PSA), which is the RRG name for a grammatical relation (Section 4). The morphosyntactic phase of the linking ends here.

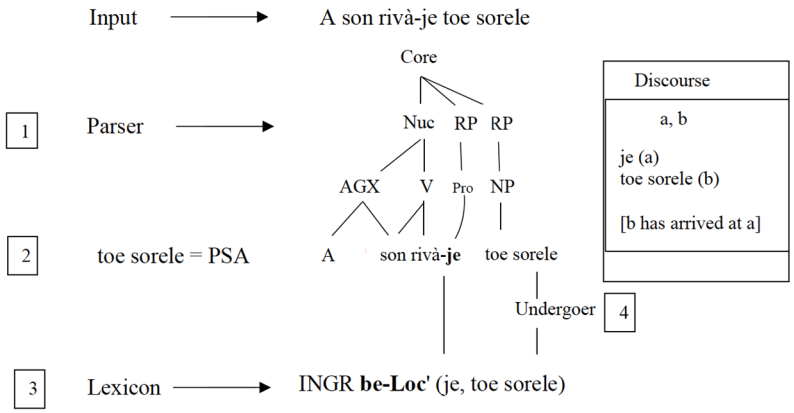


Figure 3. Pattern (iii) in syntax–semantics linking: locative interpretation of *je* in (37).

The lexical phase begins with the retrieval from the lexicon of the Logical Structure of the verb, i.e., the meaning representation with which it is stored. The RPs that have been introduced in discourse are assigned argument positions in this Logical Structure, following general lexical–decompositional principles (Van Valin and LaPolla 1997, pp. 113–16) (Step 3). The clitic *je* could in principle resume a locative adverbial that has been introduced into discourse previously or introduce anew a locative argument of the verb. Given that this is assumed to be an out-of-the blue utterance, *je* receives the latter interpretation. This is

represented in the Discourse Representation Structure, which parallels the syntax-semantics linking in Figure 3 (see the box called Discourse). This structure includes no presupposition and introduces anew the two variables a and b, their values *je* and *toe sorele*, and the semantic relation established between them in the proposition. In the final step (Step 4), the only direct core argument of the verb, *toe sorele* 'your sisters', is assigned the macrorole undergoer.[12] The sentence has been interpreted in full.

Starting from the consideration that the structure under discussion is a presentational construction, i.e., a construction which introduces an event into the universe of discourse (Section 1), and which does not provide an argument that can serve as the aboutness topic of the predication, we claim that another construal of the clitic *je* is possible. This interpretation is different from the locative reading in a subtle but significant way: *je* is not interpreted as the spell-out of the goal location of the event of arrival, but rather as a marker of the deixis of the discourse situation in which the event is announced and to which the event has relevance. The locative meaning of *je* plays a key role in this interpretation because the deixis of the discourse situation includes the spatio–temporal coordinates of the announcement. The locative deixis of the discourse situation and the goal location of the arrival can, of course, coincide. We should also add that the deixis of the discourse situation includes the speaker, who is the deictic centre of discourse (Vanelli 1972). Indeed, *je* can also indicate the spatio-temporal coordinates of the speaker at the time of the event, the speaker having experienced the event first-hand and providing a direct connection between the event and the discourse situation.

Adopting traditional terminology, in this construal *je* can be said to be—or to mark—the subject of predication. From our perspective, however, a subject of predication need not be a syntactic subject (or PSA). Indeed, *je* is not a syntactic subject in (37): it does not occur in a subject clitic position, nor does the inflected verb agree with it, as it agrees with the postverbal NP instead. In our analysis, the subject of predication is an aboutness topic, i.e., what the sentence, or utterance, is about (Gundel 1988). With specific respect to *je*, we claim that it is an index of the discourse situation, which is the aboutness topic of the utterance.

In Figure 4, we represent the syntax–semantics linking of (37), assuming *je* is an index of the discourse situation. The steps in the linking are the same as discussed above, with one important difference: since it encodes the spatio–temporal coordinates of the discourse situation, *je* is linked to the Discourse Representation Structure, which introduces these coordinates. The linking with the semantics of the verb is possible but not necessary, which explains why reanalysis can occur and why the syntax–semantics linking can be lost in diachrony. Note that the postverbal RP is interpreted as the PSA in Figure 4, i.e., the controller of verbal agreement, a grammatical relation which, in our framework, need not correlate with a specific position in syntax. This is in line with the theoretical assumptions on grammatical relations which are independently made in our framework (Section 4).

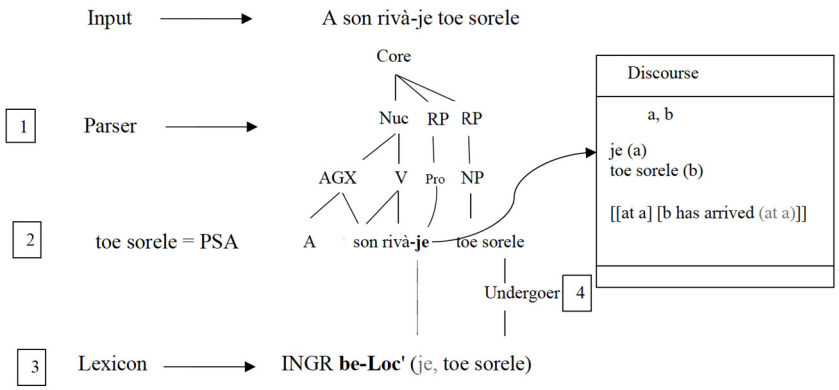


Figure 4. Pattern (iii) in syntax–semantics linking: interpretation of *je* in (37) as an index of the discourse situation.

It is this construal of *je* that can be extended to presentational constructions with verbs which do not describe motion. A relevant example is given in (38) (cf. 33b).

(38)	**A**	**son**	dimagrì**je**	tanti	cit.	(Turinese)
	SCL.3PL	be.3PL	lose.weight.PSTP.PRESCL	many	children	
	'Many children have lost weight.'					

The logical structure of *dimagrì* 'lose weight' does not comprise a locative argument. Therefore, unless a location is in the presupposition, in which case *je* will be a resumptive clitic, but (38) will not be an all-new utterance, *je* can only be understood as the marker of the deixis of the discourse situation in which the event is announced. We represent the latter alternative in Figure 5. The key point to note here is that *je* is solely interpreted through a direct linking from syntax to discourse. Thus, different interfaces are relevant to the interpretation of (38) compared with (37).

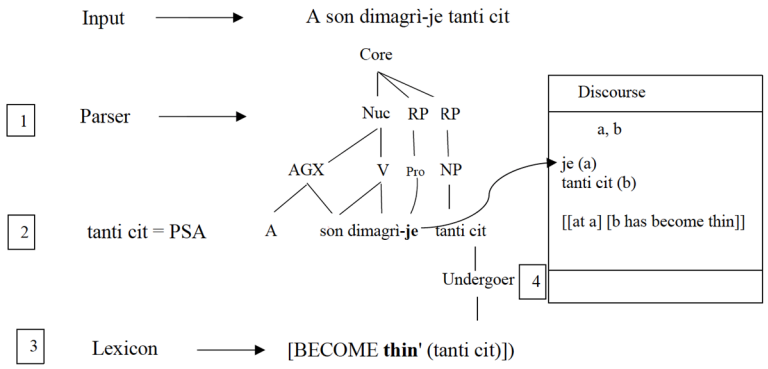


Figure 5. Pattern (iii) in syntax–semantics linking: interpretation of *je* in (38) as an index of the discourse situation.

The direct syntax–discourse linking ensures that the Completeness Constraint is satisfied and that the input is interpreted in full. While there is a notion of an Extended Completeness Constraint in RRG (Van Valin and Latrouite 2023) (Section 4), this notion has to date only been applied to cases of direct semantics–discourse, or discourse–semantics, linking, i.e., cases whereby predicates or arguments that are part of the semantics of the clause do not show up in its syntax (see, e.g., Shimojo 2008). In this article, we are applying the said notion to a direct linking which leaves out the semantic representation that is built from the Logical Structure of the verb (see Step 3 in Figure 5). This is no trivial matter in RRG, since in this framework meaning is assumed to be built compositionally, starting from the Logical Structure(s) of the predicator(s). It is therefore important to note that the insight that we are seeking to capture is not that meaning is built from syntax. Rather, our claim is that although meaning is built from the composition of Logical Structures that are stored in the Lexicon, utterances are fully interpreted within their discourse context. Following the view that there cannot be a topicless proposition, we assume that if a topic is not provided by the predicators in the clause, it will be provided by the discourse situation. The parallel architecture of RRG allows us to locate this topic where it belongs, i.e., in Discourse Representation.

5.2. Pattern (iii) vis-à-vis Patterns (i) and (ii)

As was mentioned, we take the pattern discussed in the previous section to have originated from pattern (ii), which exhibits V–S agreement but no presentational clitic. Indeed, neither pattern (ii) nor pattern (iii) were traditionally attested in Turinese (Burzio 1986; Parry 2013, among others). Pattern (ii) was then introduced because of pressure from Italian (Ricca 2008; Flecchia 2022), while pattern (iii) is the most recent one and has so far gone unnoticed. The comparative evidence from Milanese supports the hypothesis that

pattern (iii) derives from pattern (ii), since this dialect testifies to the stage that precedes the introduction of pattern (iii) into the system (see Table 2).

Let us briefly illustrate the syntax–semantics linking in patterns (i) and (ii). The linking in the former pattern is shown in Figure 6, which represents example (39) (cf. 23b).

(39)	**A**	**l'è**	nassù**je**	tante	fior.	(Turinese)
	SCL.3SG	AUXCL-be.3SG	be.born.PSTP.PRESCL	many	flowers	
	'Many flowers have appeared.'					

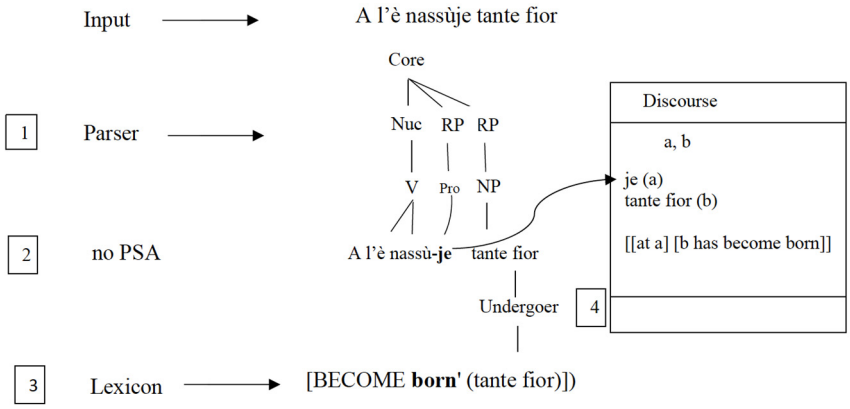


Figure 6. Pattern (i) in syntax–semantics linking: interpretation of *je* in (39) as an index of the discourse situation.

Not only is the auxiliary *è* 'is' unambiguously singular in Turinese, but *a* can be a third-person singular clitic (see note 4). The mismatch in agreement specifications between the auxiliary and the postverbal phrase is indicated by the absence of an agreement index (AGX) node in Figure 6. This mismatch results in the failure of PSA assignment, which is a key difference between the structure in (39) (Figure 6) and the one in (38) (Figure 5). The requirement of a topic is, however, satisfied by the direct syntax–discourse linking.

In our framework, the requirement of a presentational clitic in pattern (i), which is prevalent in Milanese and conservative in Turinese, can be considered to be an instruction of the presentational construction of these dialect varieties. The fact that these are subject–clitic dialects is relevant here. While in SV(O) topic-comment constructions, aboutness is expressed by S in concomitance with an agreeing subject clitic, depending on grammatical person, in VS presentationals, aboutness is expressed by a clitic alone. The autochthonous presentational construction thus has features that reflect a broader property of the grammar of these dialects.

The reverse situation is found in pattern (ii), illustrated in (40) and Figure 7. Here, a PSA is individuated, since *son* 'are' is unambiguously plural, like the postverbal phrase, and *a* can be interpreted as a third-person plural clitic, but the requirement of an aboutness topic is not satisfied overtly.

(40)	**A**		**son**	dimagrì	tuti	i	cit	(Turinese)
	SCL.3PL		be.3PL	lose.weight.PSTP	all	the	children	
	ëd	la	scola.					
	of	the	school					
	'All the children of the school have lost weight.'							

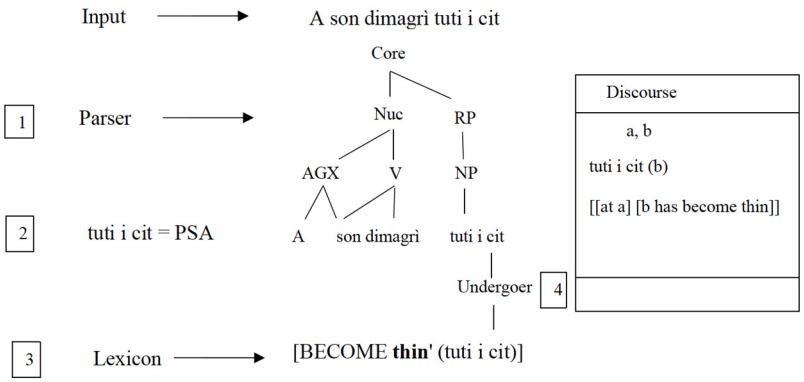


Figure 7. Pattern (ii) in syntax–semantics linking (cf. 40).

The postverbal RP is modified by the universal quantifier *tuti* 'all' in (40), and the restrictor *ëd la scòla* 'of the school' provides a frame (Lambrecht 1994, p. 90) for the interpretation of *tuti i cit* 'all the kids', whose referent is thus accessible to the interlocutors.[13] Indeed, we found that pattern (ii) was normally only deemed to be felicitous with definite postverbal noun phrases in Turinese. We should note that the same linking pattern is, however, in principle acceptable with indefinites, as can be seen in the Italian example in (41).[14]

(41)	Sono	dimagriti	tanti	bambini.	(Italian)
	be.3PL	lose.weight.PTCP	many	children	
	'Many children have lost weight.'				

In sum, the patterns analysed above demonstrate that the control verbal agreement, in our terms PSAhood, and the overt expression of an aboutness topic are in principle independent from each other. Pattern (i) is characterized by the latter but not the former (see Figure 6); in pattern (ii), we have a PSA, but not an overt expression of aboutness (see Figure 7); lastly, in pattern (iii), both aboutness and PSAhood are realized overtly, albeit separately (see Figures 4 and 5).

5.3. Subject of Predication vis-à-vis Subject

The analysis proposed in previous sections sheds light on the notion of subject of predication, treating it as orthogonal to that of controller of verbal agreement. The idea that the aboutness requirements on the clause are in principle separate from case and agreement, which, in other frameworks, can be satisfied within the verb phrase, is by no means new (Bianchi 1993; Cardinaletti 2004; Ojea 2017, etc.). In the past, however, the aboutness features of the clause have been associated with a subject position, regardless of the framework adopted in the analysis or the languages studied (Aissen 1999; Saccon 1993; Tortora 1997, 2014; Cardinaletti 2004; Parry 2013). In syntactic research, the debate has centred around the issue of whether and how this position is activated in all-new constructions (Cardinaletti 2004; Ojea 2017).

Couching our analysis in a parallel architecture framework, in this article we propose a change of perspective. We claim that, *qua* aboutness topic, the subject of predication can be an argument of the verb, which figures in the semantics and the syntax of the clause. However, it need not be an argument of the verb, in which case it is provided by the discourse situation and must be represented in discourse alone. The structures in which the subject of predication is an argument of the verb differ from those in which it is not in terms of the interfaces that are relevant to their interpretation: in the former type of structure, the subject of predication is interpreted through a direct linking between syntax and semantics (see Figure 3), which is not found in the latter type of structure (see Figures 5 and 6).

Against the backdrop of this analysis, pattern (i), which has an overt manifestation of the subject of predication, contrasts with pattern (ii), which does not, in terms of whether a linking from syntax to Discourse Representation is established in the interpretation of the sentence. The fact that only some of the languages of Italy require or allow this linking

is related to the expression of subjecthood in these languages, as has been argued in the relevant literature (Parry 2013) and as we pointed out in Section 5.2. However, we argue that this does not mean that the subject of predication is a subject. Rather, similarly to the spelling out of verbal agreement with a subject clitic, aboutness can also be marked by a clitic. Drawing on Bresnan and Mchombo (1987), Bentley (2018) suggested that this clitic expresses an anaphoric type of agreement, which is not internal to the clause or the verb phrase, but rather established between an anaphora and its antecedent in discourse. In fact, Bentley (2018) went as far as to claim that in presentational construction anaphoric agreement is in competition with the grammatical agreement between the verb and its argument. Whereas it now turns out, on evidence from contemporary Turinese (i.e., pattern (iii)), that anaphoric and grammatical agreement can both be marked overtly in presentational constructions, the connection between the morphological properties of the sister languages and the overt marking of aboutness is corroborated.

By way of conclusion of this section, we should mention and refute an alternative analysis of the data discussed in this article. It is in principle conceivable that *je* and cognates are mere markers of theticity, or of the presentational construction, without this being characterized by an aboutness topic of any sort.[15] If this were the case, the patterns discussed above would simply differ in the overt marking of theticity, while the development of the clitic would amount to a reanalysis from a locative pronoun to a theticity marker. Within the theoretical perspective of RRG, this analysis has a great deal of appeal, in that it does not require the postulation of a component of syntactic tree structure which has no correlate in semantics (see the arboreal representations in Figures 4–6) or the concomitant extension of the Completeness Constraint to cases of direct syntax–discourse linking. In addition, this analysis of *je* and cognates abides by Lambrecht's (1994) notion of sentence focus, traditionally adopted in RRG (see, e.g., Van Valin and LaPolla 1997, p. 207), which rules out a topical component.

Yet, the proposal expounded in this article is preferable on theoretical and empirical grounds. Applying Erteschik-Shir's (1997) insight that a predication is a function that maps a proposition to a topic, we rule out the notion that the interpretation of an utterance which occurs out of the blue should not start from an understanding of what the utterance is about. We claim instead that, in the absence of other clues, the utterance is understood to be about the discourse situation. Of course, the latter need not be encoded in syntax: this conclusion emerges from comparison of pattern (ii), which is the only pattern available in Italian, with patterns (i) and (iii), which we found in the dialects.

Our analysis does not violate any of the tenets of the RRG framework because the presentational clitic that is the output of the reanalysis discussed in Sections 3.2 and 5.1 is referential, its reference being in discourse and not in semantic representation. We should add that the most advanced RRG treatments of information structure do acknowledge that sentence focus is structured in a topic–comment articulation (Bentley 2023b). Therefore, our proposal constitutes another step in a direction which is already pursued within the framework.

As for the empirical advantages of the analysis proposed in this article, not only does it shed light on the microvariation attested within and across dialects, relating such variation to broader properties of the grammars of these dialects, but it can also serve as the starting point of a comparative analysis of presentationals and existentials. As was mentioned, in Italian and in some dialects of Italy, an etymologically locative clitic occurs in existentials (cf. 16) but not in presentationals (cf. 18). Since existentials are normally thetic, similarly to presentationals, this mismatch in the occurrence of the etymologically locative clitic suggests that theticity alone cannot capture the occurrence of the clitic. Further comparison of existentials and presentationals is thus needed, although it would obviously go beyond the scope of this article.

We thus propose an account of the etymologically locative clitic as a spell-out of the deixis of the discourse situation in which the presentational construction is interpreted, and we claim that the discourse situation is the aboutness topic of the all-new utterance. In

the spirit of RRG, we represent the discourse situation in Discourse Representation and we link the clitic directly from syntax to discourse.

6. Conclusions

Presentational constructions, i.e., constructions which introduce an event into the universe of discourse, raise the question of what it means for a predication to be new in information structural terms. Following a philosophical tradition established by Brentano and Marty (see note 15), such constructions have traditionally been thought to lack a topic (see Lambrecht's 1994 notion of sentence focus). However, some scholars have claimed that all predications require a subject (Bianchi 1993) or topic (Erteschik-Shir 1997), which, in the absence of co-textual or contextual clues, can be an understood, and usually speaker-oriented, spatio-temporal dimension (Benincà 1988; Calabrese 1992; Saccon 1992; see also Erteschik-Shir 1997).

The dialects of Northern Italy have figured prominently in this debate because many such dialects exhibit an etymologically locative clitic in presentational constructions, which clusters, or occurs in complementary distribution, with the subject clitics figuring in SV order. The presence of the etymologically locative clitic has traditionally been known to be mutually exclusive with verbal agreement with the postverbal argument, which thus fails to behave as a subject. Therefore, it has been claimed that the etymologically locative clitic of the presentational construction is a subject of predication, i.e., the marker of a spatio-temporal location provided by discourse and, at the same time, a subject agreement marker, comparable to the subject clitic of SV order (see, among others, Tortora 1997, 2014; Parry 2013).

In this article, we have brought to light primary evidence from Milanese and Turinese, two dialects spoken in the Northern Italian regions of Lombardy and Piedmont, respectively. We have noted that the variation attested in contemporary Turinese is more complex than has so far been noted in the literature, in that the etymologically locative clitic does not occur in a subject clitic position and is not mutually exclusive with V–S agreement, thus challenging the idea that it occurs in a subject position and marks agreement with a syntactic subject. The same evidence does not, however, challenge the view that this presentational clitic is the marker of a discourse aboutness topic.

We couched our analysis in a parallel architecture theory, Role and Reference Grammar (RRG), where the various modules of grammar (syntax, semantics, discourse) are represented independently of each other and can interact directly with each other. In our account, the presentational clitic can be a locative argument in the semantic representation of the predicate or the marker of the deixis of the discourse situation, which is an aboutness topic represented in discourse representation. In the latter case, the presentational clitic can be called a subject of predication. Crucially, our notion of subject of predication is orthogonal to that of syntactic subject and need not coincide with it. We have claimed that the disentanglement of the notions of subject of predication and syntactic subject does justice to the microvariation attested in the dialects of Northern Italy, which concerns the interfaces that are involved in the parsing of utterances and sheds light on the presentational construction itself, laying the foundations for a proper characterisation of its similarities and differences with existential constructions.

Importantly, the analysis pursued in this article considers discourse to be an integral part of grammar, an idea which has always been defended in RRG (Van Valin and LaPolla 1997; Bentley 2023b), and which is also shared by work of other theoretical persuasions (Lambrecht 1994; Erteschik-Shir 1997; Ojea 2017). Discourse is directly involved in the interpretation of utterances and in patterns of variation and processes of change, as evidenced by the reanalysis and consequent re-grammaticalization of the locative clitic of the dialects under scrutiny into a presentational clitic, which is a change in the interfaces that are relevant to the interpretation of the construction.

Author Contributions: Conceptualization, D.B. and F.M.C.; Data curation, D.B. and F.M.C.; Formal analysis, D.B.; Investigation, D.B. and F.M.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Questionnaire transcriptions are available from the authors upon request.

Conflicts of Interest: The authors declare no conflict of interest.

Notes

1 The dialects of Italy are Romance languages, and hence daughters of Latin and not varieties of Italian, the major Romance language spoken in Italy. They are conventionally referred to as *dialects* because they have very little, if any, socio-political recognition. For further detail, including the classification of these languages into different subfamilies, we refer to Parry (1997) and Loporcaro ([2013] 2020).

2 In the glosses of the examples we use the Leipzig abbreviations, with the following additions: AUXCL = auxiliary clitic; DATCL = dative clitic; EXPL = expletive; LCL = locative clitic; OCL = object clitic; PARTCL = partitive clitic; PF = (existential) proform; PRESCL = presentational clitic; PSTP = past participle; SCL = subject clitic. We maintain the original glosses of the examples that are drawn from the secondary literature.

3 Subject clitics are found in northern Italian dialects and cannot indiscriminately be assumed to be subject agreement markers (Renzi and Vanelli 1983; Rizzi 1986; Brandi and Cordin 1989; Benincà 1983, 1994; Poletto 1993, 2000; Vanelli 1997; Cardinaletti and Repetti 2010; Poletto and Tortora 2016). It would, however, go beyond the scope of this article to consider the variation in subject clitics that occurs outside the presentational construction.

4 The Turinese form *a* is a third-person singular or plural subject clitic (Regis 2006a, 2006b; Tosco et al. 2023, pp. 177–79; Regis and Rivoira 2023, p. 43). We assume that it is singular in (3) and plural in (4), in accordance with the number agreement specifications on the perfect auxiliary *esse* 'be'. We should mention that in some Northern Italian dialects there is another *a* clitic, which characterizes presentational constructions and behaves differently from subject clitics (Benincà 1983; Bernini 2012; Vai 2020). A comparative analysis of this *a* and the *a* that marks lack of number agreement in Turinese (cf. 3) is desirable but beyond the scope of this work. Here, we follow Tosco et al. (2023, p. 184) in analysing Turinese *a* as a third-person clitic, including when it occurs in presentational constructions. As for the form *l'*, it is a dummy proclitic, required by the vowel-initial forms of 'have'/ 'be' (Brandi and Cordin 1981; Pescarini 2016, pp. 748–49; Tosco et al. 2023, pp. 261–62; Regis and Rivoira 2023, p. 55). Following a long-established tradition, we gloss it as AUXCL (auxiliary clitic), regardless of whether it precedes an auxiliary or a copula.

5 The questionnaire included 36 multiple-choice dialect entries, each preceded by contextual information. The interviews were conducted in two different stages. Author A interviewed two native speakers of Milanese in the period between November 2014 and June 2015 (see Author A XXX), while Author B interviewed nine Turinese speakers in the period between December 2022 and September 2023. The native speaker informants (five women and six men) were aged between 40 and 80 years. Their level of education ranged from *scuola media* 'middle school' to *scuola superiore* 'high school', with one exception: one of the Milanese informants had completed a university degree. They were all individuals who speak the dialect on a daily basis in informal contexts, that is, with family and friends. Unless otherwise stated, the examples that we will provide illustrate the one option (out of those given as multiple choices) that was selected as the preferred choice by all the speakers of the given dialect. While a larger and numerically balanced speaker sample would have been preferable, we note that speaker numbers are low for the two dialects under investigation, particularly in the city of Milan, where the first round of interviews was conducted. It is of course possible that the apparent homogeneity of the Milanese data is a mere side-effect of the small size of the sample. Nonetheless, this has no consequences for our analysis, which does not adopt quantitative methods or aim to capture each dialect exhaustively, but rather proposes an explanation of the microvariation that we attested.

6 We should note that, in the examples with the third-person plural pronoun, verb agreement and/or lack thereof are both deemed to be acceptable by some speakers.

7 *A* is glossed as EXPL(etive) in (19) and (21) to follow Parry (2013).

8 Again, we follow Parry's (2013) glossing conventions here.

9 A note on the early varieties of the Centre-South is in order. These were null-subject vernaculars and never lacked V–S agreement. Yet, the clitic also emerged and established itself as a component of the existential construction in these vernaculars. The view that the clitic was reanalysed as a subject agreement marker satisfying a syntactic subjecthood requirement does not capture its development into an existential proform in these vernaculars. Instead, evidence from a geo-linguistically varied corpus of early Italo-Romance texts suggests that the clitic appeared in VS copular structures to resume a distant topical locative phrase. This ensured that the conditions of discourse coherence and cohesion were met in the narrative. In existentials, the clitic became the marker of the implicit contextual domain of these constructions (Francez 2007). This view of the emergence of the existential

proform (see Ciconte in Bentley et al. 2015, pp. 248–49, 254–56) accounts for the variation in V–S agreement in all the early Italo-Romance varieties but does not conflict with the analysis of the clitic as an agreement marker in the northern vernaculars, where it came to be associated with a subject position.

[10] It is worth pointing out that in the early sources presentational VS constructions are consistently introduced by spatio–temporal adverbials (e.g., *allora* 'then', *adunc(a)* 'then/at that point', *donde* 'thereafter'/'therefore', etc.) derived from locative etyma (Ciconte 2018, pp. 141–42). In the logo-deixis of the written domain, where there cannot be an implicit reference to the communicative situation, these adverbials spell out a narrative aboutness topic, similarly to how *je* spells out a discourse aboutness topic in modern Turinese.

[11] It is worth pointing out that this constructional requirement is not valid in all dialects, as testified by the dialect of Grosio, where we found a fourth pattern without clitic or V–S agreement ({-Cl; -Agr}, cf. 18a) in a previous survey (Bentley 2018).

[12] Core arguments are arguments that are required by the Logical Structure of the verb, i.e., the semantic representation that is stored in the lexicon for the verb. Direct core arguments are unmarked, or marked by case alone, differently from oblique arguments, which are adpositionally marked.

[13] These aspects of the semantics of the noun phrase would be taken care of in its semantic representation, which we do not provide here for brevity.

[14] The definiteness contrast between Turinese and Italian reflects the microvariation in the constraints on PSAhood that are at work in the two languages, an issue which goes beyond the scope of this article (see Bentley 2018; Bentley and Cennamo 2022).

[15] This hypothesis was suggested to us by Jürgen Bohnemeyer at the 17th International Conference on Role and Reference Grammar (Heinrich Heine University of Düsseldorf, 14–16 August 2023). We refer the reader to Kuroda (1972) and Sasse (1987) for the thetic/categorical distinction, which originated in the work of the Swiss philosopher of language Anton Marty (1847–1914), who in turn developed ideas by the German philosopher Franz Clemens Brentano (1838–1917).

References

Aissen, Judith. 1999. Markedness and subject choice in Optimality Theory. *Natural Language and Linguistic Theory* 17: 673–711. [CrossRef]

Benincà, Paola. 1983. Il clitico «a» nel dialetto padovano. In *Scritti Linguistici in Onore di G.B. Pellegrini*. Edited by Paola Benincà and Laura Vanelli. Pisa: Pacini, pp. 25–35, Reprinted in 1994.

Benincà, Paola. 1988. L'ordine degli elementi della frase: Costruzioni con ordine marcato degli elementi. In *Grande Grammatica Italiana di Consultazione*. Edited by Renzi Lorenzo. Bologna: Il Mulino, vol. 1, pp. 129–94.

Benincà, Paola. 1994. Il clitico 'a' nel dialetto padovano. In *La Variazione Sintattica: Studi di Dialettologia Romanza*. Edited by Benincà Paola. Bologna: Il Mulino, pp. 15–27, First publish 1983.

Benincà, Paola. 2007. Clitici e ausiliari: Gh ò, z è. In *Sui Dialetti Italoromanzi: Saggi in Onore di Nigel B. Vincent*. Edited by Delia Bentley and Ledgeway Adam. *The Italianist* 27, Special Supplement 1. King's Lynn: Biddles, pp. 27–47.

Benincà, Paola, and Laura Vanelli. 1986. Punti di sintassi comparata dei dialetti italiani settentrionali. In *Raetia Antiqua et Moderna: W. Th. Elwert zum 80. Geburtstag*. Edited by Gunter Holtus and Kurt Ringger. Tübingen: Niemeyer, pp. 457–79.

Bentley, Delia. 2018. Grammaticalization of subject agreement on evidence from Italo-Romance. *Linguistics* 56: 1245–301. [CrossRef]

Bentley, Delia. 2023a. Introduction. In *The Cambridge Handbook of Role and Reference Grammar*. Edited by Delia Bentley, Ricardo Mairal Usón, Wataru Nakamura and Robert D. Van Valin, Jr. Cambridge: Cambridge University Press, pp. 1–14.

Bentley, Delia. 2023b. The RRG approach to information structure. In *The Cambridge Handbook of Role and Reference Grammar*. Edited by Delia Bentley, Ricardo Mairal Usón, Wataru Nakamura and Robert D. Van Valin, Jr. Cambridge: Cambridge University Press, pp. 457–87.

Bentley, Delia, and Michela Cennamo. 2022. Thematic and lexico-aspectual constraints on V-S agreement: Evidence from Northern Italo-Romance. In *Periphrasis and Inflexion in Diachrony: A View from Romance*. Edited by Adam Ledgeway, John Charles Smith and Nigel Vincent. Oxford: Oxford University Press, pp. 335–61.

Bentley, Delia, Francesco Maria Ciconte, and Silvio Cruschina. 2015. *Existentials and Locatives in Romance Dialects of Italy*. Oxford: Oxford University Press.

Bentley, Delia, Ricardo Mairal Usón, Wataru Nakamura, and Robert D. Van Valin, Jr. 2023. *The Cambridge Handbook of Role and Reference Grammar*. Cambridge: Cambridge University Press.

Bernini, Giuliano. 2012. Il clitico a nell'italo-romanzo settentrionale: Osservazioni metodologiche. In *Per Roberto Gusmani: Linguistica Storica e Teorica*. Edited by Vincenzo Orioles. Udine: Forum, vol. 2, pp. 269–82.

Bianchi, Valentina. 1993. Subject positions and e-positions. In *Quaderni del laboratorio di linguistica 7*. Pisa: Scuola Normale Superiore, pp. 51–69.

Brandi, Luciana, and Patrizia Cordin. 1981. Dialetti e italiano: Un confronto sul parametro del soggetto nullo. *Rivista di Grammatica Generativa* 6: 33–87.

Brandi, Luciana, and Patrizia Cordin. 1989. Two Italian dialects and the null subject parameter. In *The Null Subject Parameter*. Edited by Osvaldo Jaeggli and Kenneth J. Safir. Dordrecht: Kluwer, pp. 111–42.

Bresnan, Joan, and Sam A. Mchombo. 1987. Topic, pronoun and agreement in Chichewa. *Language* 63: 741–82. [CrossRef]

Burzio, Luigi. 1986. *Italian Syntax. A Government–Binding Approach*. Dordrecht: Kluwer.

Calabrese, Andrea. 1992. Some remarks on focus and logical structure in Italian. *Harvard Working Papers in Linguistics* 1: 91–127.
Cardinaletti, Anna. 2004. Toward a cartography of subject positions. In *The Structure of CP and IP. The Cartography of Syntactic Structures*. Edited by Luigi Rizzi. New York: Oxford University Press, vol. 2, pp. 115–65.
Cardinaletti, Anna, and Lori Repetti. 2010. Proclitic vs enclitic pronouns in northern Italian dialects and the null-subject parameter. In *Syntactic Variation. The Dialects of Italy*. Edited by Roberta D'Alessandro, Adam Ledgeway and Ian Roberts. Cambridge: Cambridge University Press, pp. 119–34.
Ciconte, Francesco Maria. 2008. Existential constructions in early Italo-Romance vernaculars. In *Proceedings of the Oxford Postgraduate Conference LingO 2007*. Edited by Miltiadis Kokkonidis. Oxford: University of Oxford, pp. 35–42.
Ciconte, Francesco Maria. 2011. The emergence and the reanalysis of the existential proform: Evidence from early Italo-Romance. *Transactions of the Philological Society* 109: 284–306. [CrossRef]
Ciconte, Francesco Maria. 2013. Argument realization and existential proforms in early Italo-Romance. In *Argument Structure in Flux: The Naples/Capri Papers*. Edited by Jóhanna Barðdal, Michela Cennamo and Elly van Gelderen. Amsterdam: Benjamins, pp. 549–66.
Ciconte, Francesco Maria. 2018. Postverbal subjects in old Italo-Romance. *Italian Journal of Linguistics* 30: 127–58.
Erteschik-Shir, Nomi. 1997. *The Dynamics of Focus Structure*. Cambridge: Cambridge University Press.
Flecchia, Alessandro. 2021. Fenomeni di Italianizzazione in Alcuni Dialetti di Area Biellese. Master's thesis, University of Turin, Turin, Italy.
Flecchia, Alessandro. 2022. Fenomeni di italianizzazione in alcuni dialetti biellesi. *Bollettino dell'Atlante Linguistico Italiano* III: 33–64.
Foley, William, and Robert D. Van Valin, Jr. 1984. *Functional Syntax and Universal Grammar*. Cambridge: Cambridge University Press.
Francez, Itamar. 2007. Existential Propositions. Ph.D. thesis, Stanford University, Stanford, CA, USA.
Gundel, Jeanette. 1988. *The Role of Topic and Comment in Linguistic Theory*. New York: Garland.
Hopper, Paul J., and Elizabeth C. Traugott. 1993. *Grammaticalization*. Cambridge: Cambridge University Press.
Jackendoff, Ray. 2002. *Foundations of Language: Brain, Meaning, Grammar, Evolution*. Oxford: Oxford University Press.
Kamp, Hans, and Uwe Reyle. 1993. *From Discourse to Logic*. Hingham: Kluwer.
Karssenberg, Lena. 2016. French 'il y a' clefts, existential sentences and the focus-marking hypothesis. *Journal of French Language Studies* 27: 1–26.
Karssenberg, Lena. 2018a. *Non-Prototypical Clefts in French: A Corpus Analysis of il y a Clefts*. Berlin and Boston: De Gruyter.
Karssenberg, Lena. 2018b. Non-prototypical clefts in French. A corpus analysis of il y a clefts. In *Beihefte zur Zeitschrift für romanische Philologie*. Band 424. Berlin: De Gruyter.
Kuroda, Sige-Yuki. 1972. The categorical and the thetic judgment: Evidence from Japanese Syntax. *Foundations of Language* 9: 153–85.
Lambrecht, Knud. 1988. Presentational cleft constructions in spoken French. In *Clause Combining in Grammar and Discourse*. Edited by John Haiman and Sandra A. Thompson. Amsterdam and Philadelphia: John Benjamins Publishing Company, pp. 135–79.
Lambrecht, Knud. 1994. *Information Structure and Sentence Form: Topic, Focus, and the Mental Representation of Discourse Referents*. Cambridge: Cambridge University Press.
LaPolla, Randy. 2023. Grammatical Relations. In *The Cambridge Handbook of Role and Reference Grammar*. Edited by Delia Bentley, Ricardo Mairal Usón, Wataru Nakamura and Robert D. Van Valin, Jr. Cambridge: Cambridge University Press, pp. 269–91.
Loporcaro, Michele. 2020. *Profilo linguistico dei dialetti italiani*. Bari: Laterza. First published 2013.
Manzini, Maria Rita, and Leonardo Savoia. 2005. *I Dialetti Italiani e Romanci: Morfosintassi Generativa*. Alessandria: Edizioni Dell'Orso, vol. 2.
Ojea, Ana. 2017. Core intentional features in the syntactic computation: Deriving the position of the subject in Spanish. *Lingua* 195: 72–91. [CrossRef]
Parry, Mair. 1997. Piedmont. In *The Dialects of Italy*. Edited by Martin Maiden and Mair Parry. London: Routledge, pp. 237–44.
Parry, Mair. 2000. Accordo e soggetti postverbali in piemontese. In *Actes du XXIIe Congrès International de Lingüistique e Philologie Romane, Bruxelles 1998, VI. De la Grammaire des Formes à la Grammaire du Sens*. Tübingen: Niemeyer, pp. 391–402.
Parry, Mair. 2010. Non-canonical subjects in the early Italian vernaculars. *Archivio Glottologico Italiano* 95: 190–226.
Parry, Mair. 2013. Variation and change in the presentational constructions of north-western Italo-Romance varieties. In *Argument Structure in Flux: The Naples/Capri Papers*. Edited by Jóhanna Barðdal, Michela Cennamo and Elly van Gelderen. Amsterdam: Benjamins, pp. 511–48.
Pescarini, Diego. 2016. Clitic pronominal systems: Morphophonology. In *The Oxford Guide to the Romance Languages*. Edited by Adam Ledgeway and Martin Maiden. Oxford: Oxford University Press, pp. 742–57.
Poletto, Cecilia. 1993. *La sintassi del Soggetto nei Dialetti Italiani Settentrionali*. Padua: Unipress.
Poletto, Cecilia. 2000. *The Higher Functional Field. Evidence from Northern Italian Dialects*. Oxford: Oxford University Press.
Poletto, Cecilia, and Christina Tortora. 2016. Subject clitics: Syntax. In *Oxford Guide to the Romance Languages*. Edited by Adam Ledgeway and Martin Maiden. Oxford: University Press, pp. 772–85.
Regis, Riccardo. 2006a. I pronomi clitici soggetto nel Piemonte occidentale. *Lingue e Idiomi d'Italia* 1: 53–85.
Regis, Riccardo. 2006b. Se i clitici soggetto sono facoltativi: Il caso piemontese. *Studi e Ricerche Quaderni del Dipartimento di Scienze del Linguaggio e Letterature Moderne e Comparate dell'Università di Torino* 1: 191–232.
Regis, Riccardo, and Matteo Rivoira. 2023. *Dialetti d'Italia. Piemonte e Valle d'Aosta*. Roma: Carocci.

Renzi, Lorenzo, and Laura Vanelli. 1983. I pronomi soggetto in alcune varietà romanze. In *Scritti Linguistici in Onore di Giovan Battista Pellegrini*. Edited by Paola Benincà, Manlio Cortelazzo, Aldo Prosdocimi, Laura Vanelli and Alberto Zamboni. Pisa: Pacini, pp. 121–45.

Ricca, Davide. 2008. Tratti instabili nella sintassi del piemontese contemporaneo: Tra italianizzazione e arcaismi locali. In *Sprachwandel und (Dis-)Kontinuität in der Romania*. Edited by Sabine Heinemann. Tubingen: Niemeyer, pp. 113–27.

Rizzi, Luigi. 1986. On the status of subject clitics in Romance. In *Studies in Romance Linguistics*. Edited by Osvaldo Jaeggli and Carmen Silva-Corvalan. Dordrecht: Foris, pp. 391–419.

Roberts, Ian. 1993. Inversion and subject clitics in Valdôtain. *Edinburgh Working Papers in Cognitive Science* 6: 155–67.

Saccon, Graziella. 1992. VP-internal arguments and locative subjects. In *Proceedings of the 22nd Meeting of the North Eastern Linguistic Society*. Amherst: GLSA, UMass/Amherst, pp. 383–97.

Saccon, Graziella. 1993. Post-Verbal Subjects: A Study Based on Italian and Its Dialects. Ph.D. thesis, Harvard University, Cambridge, MA, USA.

Sasse, Hans-Jürgen. 1987. The thetic/categorical distinction revisited. *Linguistics* 25: 511–80. [CrossRef]

Shimojo, Mitsuaki. 2008. How missing is the missing verb? The verb-less numeral quantifier construction in Japanese. In *Investigations of the Syntax-Semantics Pragmatics Interface*. Edited by Robert D. Van Valin, Jr. Amsterdam: John Benjamins, pp. 285–304.

Tortora, Christina. 1997. The Syntax and Semantics of the Weak Locative. Ph.D. thesis, University of Delaware, Newark, DE, USA.

Tortora, Christina. 2014. *A Comparative Grammar of Borgomanerese*. New York: Oxford University Press.

Tosco, Mauro, Emanuele Miola, and Nicola Duberti. 2023. *A Grammar of Piedmontese. A Minority Language of Northwest Italy*. Leiden: Brill.

Vai, Massimo. 2020. Il clitico «a» nella storia del milanese. In *Italiani di Milano. Studi in Onore di Silvia Morgana*. Edited by Giuseppe Sergio and Massimo Prada. Milan: Ledizioni, vol. 8, pp. 59–77.

Van Valin, Robert D., Jr. 2005. *Exploring the Syntax-Semantics Interface*. Cambridge: Cambridge University Press.

Van Valin, Robert D., Jr. 2023. Principles of Role and Reference Grammar. In *The Cambridge Handbook of Role and Reference Grammar*. Edited by Delia Bentley, Ricardo Mairal Usón, Wataru Nakamura and Robert D. Van Valin, Jr. Cambridge: Cambridge University Press, pp. 17–177.

Van Valin, Robert D., Jr., and Anja Latrouite. 2023. Information Structure and Argument Linking. In *The Cambridge Handbook of Role and Reference Grammar*. Edited by Delia Bentley, Ricardo Mairal Usón, Wataru Nakamura and Robert D. Van Valin, Jr. Cambridge: Cambridge University Press, pp. 488–522.

Van Valin, Robert D., Jr., and Randy LaPolla. 1997. *Syntax. Structure, Meaning and Function*. Cambridge: Cambridge University Press.

Vanelli, Laura. 1972. *La Deissi in Italiano*. Padua: Unipress.

Vanelli, Laura. 1997. Personal pronouns and demonstratives. In *The Dialects of Italy*. Edited by Martin Maiden and Mair Parry. London: Routledge, pp. 106–15.

von Heusinger, Klaus. 1999. Intonation and Information Structure. In *The Representation of Focus in Phonology and Semantics. Habilitationsschrift*. Konstanz: University of Konstanz.

languages

MDPI

Article

Towards a Typology of Zero Aboutness: Expletive *A* in Fornese and *Chiru* in Cilentano

Simone De Cia [1,*] and Mariangela Cerullo [2]

[1] Department of Linguistics and English Language, University of Manchester, Manchester M13 9PL, UK
[2] Department of Humanities, University of Naples Federico II, 80138 Napoli, NA, Italy; mariangela.cerullo@unina.it
* Correspondence: simone.decia@manchester.ac.uk

Abstract: This paper investigates the syntactic–pragmatic behavior of two expletive-like elements, namely *a* and *chiru*, in Fornese and Cilentano, two Romance varieties spoken in Northern and Southern Italy, respectively. We argue that *a* and *chiru* are not *bona fide* expletive subjects but discourse-pragmatic expletives, which mark *zero aboutness* or the absence of an aboutness referent in an utterance. The investigation of Fornese and Cilentano points towards the existence of a sub-class of null-subject languages where aboutness as a discourse feature must be structurally satisfied by merging an overt or null topic in the syntactic spine of the clause. In the absence of such an element—for example, in thetic clauses—a discourse-pragmatic expletive is externally merged as a last-resort strategy to satisfy [uAboutness]. We argue that, in these null-subject languages, the satisfaction of the discourse feature [uAboutness] is an LF requirement, which is subject to a parametric choice. We show that, in Fornese, "default" [aboutness] is satisfied in SubjP, which is the canonical syntactic position for overt subjects within a cartographic approach. In Cilentano, on the other hand, [aboutness] is satisfied in a higher position within the C-domain, namely ShiftP, the canonical syntactic position that hosts overt aboutness/shift topics.

Keywords: aboutness/shift; expletive; Italian Dialects; morpho-syntax; topic

Citation: De Cia, Simone, and Mariangela Cerullo. 2024. Towards a Typology of Zero Aboutness: Expletive *A* in Fornese and *Chiru* in Cilentano. *Languages* 9: 60. https://doi.org/10.3390/languages9020060

Academic Editor: Ana Ojea

Received: 29 November 2023
Revised: 27 January 2024
Accepted: 28 January 2024
Published: 7 February 2024

1. Introduction

Expletive subjects are argued to be a last-resort strategy to satisfy the formal requirement of marking the canonical subject position in languages where this syntactic slot must be phonologically realized. In generative syntax, this requirement has been theoretically formalized as the satisfaction of an Extended Projection Principle (EPP) feature, which involves the lexicalization of SpecTP, or SpecSubjP within a cartographic approach (Chomsky 1995, 2004; Rizzi and Shlonsky 2006, 2007). Since the formulation of the Null-Subject Parameter (Chomsky 1981), null-subject languages have been assumed not to license overt expletives, as languages that have null referential subjects can also license null non-referential subjects (Rizzi 1982, 1986). Furthermore, it has also been argued that pro-drop languages can satisfy the EPP via alternative syntactic mechanisms, for example, via V-to-T movement (see Alexiadou and Anagnostopoulou 1998), making expletive subjects completely redundant in null-subject languages. Nevertheless, there is a growing body of literature that describes expletive-like elements in pro-drop languages (Holmberg and Nikanne 2002; Hinzelin and Kaiser 2007; Carrilho 2008; Nicolis 2008; Kaiser and Remberger 2009; Bartra-Kaufmann 2011, a.o.). These expletives appear in those syntactic environments where an expletive subject proper would appear in non-null-subject languages; nevertheless, they tend to be optional and sensitive to discourse-pragmatics, often encoding a "speaker-related" meaning (see Greco et al. 2017). This paper provides novel data on the syntactic distribution and morpho-syntactic status of two discourse-pragmatic expletives found in two null-subject Romance varieties, namely Fornese (cf. 1), spoken in the North-Eastern part of Italy, and Cilentano (cf. 2), spoken in Southern Italy.[1]

1. A ì muart las vacias
   EXPL be.3SG die.PTCP the.FPL cow.FPL
   'There died the cows.'

2. Chiru a muortu mariti e muglieri
   EXPL have.3SG die.PTCP husband.M.PL CONJ wife.FPL
   'There died husband and wife.'

At *prima facie*, Fornese *a* and Cilentano *chiru* seem to function as syntactic placeholders for the subject position. However, we will show that, in the two pro-drop languages, these expletive-like elements do not lexicalize the canonical subject position but are the manifestation of a formal requirement at the syntax–pragmatics interface. More specifically, the investigation of Fornese *a* and Cilentano *chiru* points towards the existence of a subclass of null-subject languages where the lack of an aboutness referent, be it explicit or null (i.e., presupposed), must be overtly marked in the syntactic spine of the clause. An aboutness referent can be either an overt or null aboutness/shift topic or an overt lexical or pronominal element in a preverbal position. Rizzi and Shlonsky (2006, 2007) claim, in fact, that *aboutness* can also be encoded by the canonical subject of a clause. *A* and *chiru* are two discourse-pragmatic expletives that serve the same function: to overtly syntactically mark zero aboutness (i.e., the absence of an explicit or null aboutness XP in the sentence). Our claim is in line with Erteschik-Shir's (1999) view that the truth value of the propositional content of all clauses must be checked against a *topic* (in the sense of Reinhart 1981), and hence also all-new-information sentences possess a topic-comment articulation. At the syntax–pragmatics interface, we claim that Fornese and Cilentano must satisfy this requirement structurally by saturating an [uAboutness] feature in the spine of the clause.[2] We will argue that *a* and *chiru* signal that no aboutness topic is present in the utterance (i.e., zero aboutness), and a new aboutness topic must be selected from the propositional content of the following all-new-information sentence.

The presence of an expletive element linked to the lack of topicality in the clause is not an entirely new claim in the literature, especially with reference to the Germanic languages. On the matter, Sasse (1987), Lambrecht and Polinsky (1997), and Lambrecht (2000) use terms such as *desubjectivization* and *detopicalization* to describe the various strategies languages employ to signal lack of topicality in the clause. In a diachronic perspective, Faarlund (1990) discusses the emergence of expletive topics as a means to satisfy the verb-second constraint in Germanic (see also Rögnvaldsson and Thráinsson 1990, on Icelandic). Poletto (2005) puts forward a comparable claim with respect to the topic marker *e* in Old Italian. The novelty of this paper lies in shedding light on the type of topicality that triggers this phenomenon. Topics are not a homogeneous class but rather serve different discourse-pragmatic functions. The literature of the past three decades has convincedly shown that there exist at least three types of topics, namely *aboutness/shift topics*, *given* or *familiar topics*, and *contrastive topics* (Frascarelli and Hinterhölzl 2007; Bianchi and Frascarelli 2010). These broadly correlate with three macro discourse-pragmatic functions— *aboutness*, *givenness*, and *contrastiveness*, respectively. We will show that, in Fornese and Cilentano, discourse-pragmatic expletives do not surface due to the lack of any topical element but, more specifically, due to the lack of an overt or null XP, which carries an aboutness/shift interpretation. By adopting a cartographic approach, in which discourse features are directly responsible for the discourse-pragmatic interpretation of XPs in specific syntactic configurations (Rizzi 1997; Cinque 1999; Cruschina 2012), Fornese and Cilentano lend evidence that the satisfaction of [uAboutness] in relation to zero aboutness is subject to a parametric choice within the syntactic spine of the clause. We will argue that, in Fornese, "default" aboutness is satisfied in SubjP—the canonical syntactic position for overt subjects (Rizzi and Shlonsky 2006, 2007; Bentley and Cruschina 2018), whereas in Cilentano in ShiftP—the canonical syntactic position for aboutness/shift topics (Bianchi and Frascarelli 2010).

This paper is structured as follows. In Section 2, we discuss the morpho-syntactic distribution of *a* and *chiru*, along with some methodological considerations on data collection.

In Section 3, we show that the manifestation of discourse-expletives in Fornese and Cilentano is linked to zero aboutness. Section 4 proposes two left-peripheral syntactic positions for the satisfaction of [uAboutness] and puts forward the claim that the manifestation of zero aboutness is subject to parametric variation. In light of Chomsky's (2001, 2004) Agree probe-goal model, Section 4 also provides a syntactic account of the phenomenon. Section 5 concludes the paper.

2. Fornese *A* and Cilentano *Chiru*: Overview and Morpho-Syntactic Distribution

This paper analyzes the syntactic–pragmatic behavior of two expletive-like elements in two understudied null-subject Romance languages spoken in Italy, namely Fornese *a* and Cilentano *chiru*. As far as their genetic affiliation is concerned, Fornese and Cilentano lie on opposite sides with respect to the Romance north–south divide (Zamboni 1998, drawing on La Fauci 1988; Renzi and Andreose 2015). Fornese is spoken in the North-Eastern part of Italy in the mountainous and isolated municipality of Forni di Sopra (province of Udine) by roughly one thousand speakers. Fornese shares its linguistic traits with Carnic or Northern Friulian (Benincà and Vanelli 2016) and, to a lesser extent, Cadorino Ladin (Pellegrini 1979). Cilentano is instead the name for a dialect continuum of vernacular Campanian dialects spoken in Southern Italy. More specifically, Cilentano is spoken in the area of Cilento in the province of Salerno by roughly two hundred fifty thousand speakers. As we will discuss in Sections 2.2 and 2.3, respectively, Fornese *a* is a weak pronominal element completely bleached of any lexical meaning, whereas Cilentano *chiru* is a desemantized tonic pronoun that developed from the homophonous third-person singular masculine distal demonstrative pronoun *chiru* 'that'.

Whilst it may be argued that the presence of expletive-like *a* in Fornese is linked to the hybrid status of Northern Italian Dialects (abbreviated NIDs, see ft. 1) as non-consistent null-subject languages (Cardinaletti and Repetti 2010), the status of *chiru* in Cilento is more puzzling, as Southern Italian Dialects are generally "well-behaved" null-subject languages. We will show that these expletive-like elements do not function like subject expletives proper, but lexically mark an empty aboutness/shift topic position, namely zero aboutness. By adopting a cartographic approach, we will also show that, despite the comparable syntactic distribution of *a* and *chiru*, the lexicalization of zero aboutness is subject to parametric variation, targeting different functional projections in the syntactic spine of the clause. To this aim, we will present novel data collected during several field trips to the municipalities of Forni di Sopra (province of Udine), Felitto, and Piaggine (province of Salerno). In order to maximize the naturalness of the elicited discourse-pragmatic data, interviews were carried out in small groups (roughly three groups per speech community) of three or four speakers. We gathered eighteen hours of recordings: eight hours for Fornese and ten hours for Cilentano. We heavily draw on *questionnaire-based elicitation* and *naturally occurring data* (see Himmelmann 1998, 2006; Milroy and Gordon 2003). *Data manipulation* and subsequent *acceptability judgments* were also partly used as a tool of investigation (Chelliah and de Reuse 2011). Note that all constructions that were found through elicitation were also attested in naturally occurring speech. It goes without saying that no written corpora exist of these spoken Romance languages; as a result, first-hand data collection is the only possible means to study them.

In the following sections, we will first show the striking similarities with respect to the type of syntactic environments in which Fornese *a* and Cilentano *chiru* are found as discourse-pragmatic expletives. We will then look at *a* and *chiru* separately, distinguishing their discourse-pragmatic expletive function from any other morpho-syntactic function they may have in the languages.

2.1. The Surfacing Contexts of Discourse-Pragmatic Expletives A and Chiru

The lexicalization of discourse-pragmatic expletives *a* and *chiru* occurs in two specific syntactic contexts, primarily characterized by the absence of a lexical or pronominal subject. First, *a a*nd *chiru* obligatorily surface in those syntactic environments where a non-null-

subject language like English features obligatory subject expletives, namely with weather verbs, presentational and existential constructions, impersonal clauses, and in the case of extraposition (see Williams 2001; Biberauer and Roberts 2010; Pescarini 2014). Second, *a* and *chiru* can be optionally found in a clause featuring a null referential subject, crucially in complementary distribution with an overt pronominal or lexical subject.

Let us start by exemplifying those syntactic environments where a subject expletive proper would be found in a non-null-subject language, in which the occurrence of *a* and *chiru* is obligatory.[3] As shown in (3) to (6), these discourse-pragmatic expletives *a* and *chiru* are found in presentational (cf. 3 and 5) and existential constructions (cf. 4 and 6):

Fornese

3. A ì colât i plat-s
   EXPL be.3SG fall.PTCP the plate-PL
   'There fell the plates.'

4. A era de las fantata-s in tal bosc
   EXPL be.3SG.PST of the young.woman-PL in the woods
   'There were some young women in the woods.'

Cilentano

5. Chiru è chiusu a lavanderia
   EXPL be.3SG close.PTCP the laundry
   'The launderette is closed.'

6. Chiru nge foje la pesta ccane
   EXPL PF be.PST.3SG the pest here
   'There was a pestilence here.'

The lexicalization of *a* and *chiru* patterns with the emergence of default third-person singular agreement on the inflected verb, regardless of the person and number of the plural postverbal argument (cf. 2 and 3). This is systematically found in both Fornese and Cilentano. As far as Example (4) is concerned, it is important to note that, similarly to Friulian, Fornese lacks an existential-locative proform, like *ci* in Italian or *ghe* in Venetan (see Bentley et al. 2015), which is instead present in Cilentano, namely *ngi* (<Lat. hince 'hence' < ECCE HIC, Rohlfs 2021). The presence or absence of the existential-locative particle is nonetheless orthogonal to the claims put forward in this paper.[4]

Fornese and Cilentano obligatorily also feature *a* and *chiru* with meteorological verbs, as shown in (7) and (9), and impersonal constructions, as in (8) and (10):

Fornese

7. A niviê su la tsima da-i mons
   EXPL snow.3SG on the top of-the mountains
   'It's snowing on the top of the mountains.'

8. A si dopra dapardut chesta roba achi
   EXPL IMP use.3SG everywhere this thing here
   'One uses this thing here everywhere.'

Cilentano

9. Chiru vendèa buono ra rupe
   EXPL wind.blow.3SG good from cliff
   'It is very windy on the cliff.'

10. Chiru non se pòte passà u ponde cu a Maronna
    EXPL NEG IMP can.3SG pass.INF the bridge with the Virgin.Mary
    'It is prohibited to cross the bridge carrying the statue of the Virgin Mary.'

While the presence of *a* and *chiru* in impersonal constructions is well-behaved, meteorological verbs exhibit some idiosyncrasies both in Fornese and Cilentano. We claim that

this is due to the *quasi-argumental* nature of the subject of this class of verbs. In broad terms, weather-verb expletives should be distinguished from other types of expletives, as weather verbs are argued to retain partial argument structure (Bolinger 1977; Chomsky 1981; Manzini and Savoia 2005; Levin 2015). In fact, in Cilentano, *chiru* can be commonly replaced either by the proximal demonstrative form *chistu* 'this' or by a lexical subject like *lu tjempu* 'the weather'.[5] Consequently, the third-person singular masculine distal demonstrative pronoun *chiru* may thus not be entirely non-referential with weather verbs. In this specific context, *chiru* may be seen as serving a dual purpose: (a) spell out the quasi-argument of whether predicates and (b) satisfy the aboutness of the clause. It goes without saying that if a lexical expression like *lu tiempu* is used, *chiru* cannot surface. On the other hand, Fornese does not allow a lexical subject with weather verbs; nevertheless, as we will further discuss in the next section (i.e., 2.2), weather verbs can optionally be accompanied by a third-person masculine singular subject clitic. If the subject clitic is present, we assume that the null subject of the weather verb encodes some referentiality; this, however, does not hinder the surfacing of the discourse-pragmatic expletive *a* in Fornese. In the two languages, the morpho-syntactic behavior of weather verbs in relation to discourse-pragmatic expletives *a* and *chiru* lend support to the view that the subjects of weather verbs exhibit *quasi-argumental* properties. We will not explore the matter further; nevertheless, we will partly continue the discussion in Section 3. Despite these idiosyncrasies, the lexicalization of *a* and *chiru* in this syntactic environment is robustly attested.

As for those syntactic environments canonically associated with the surfacing of subject expletives proper in non-null-subject languages, *a* and *chiru* are also found with extraposition (cf. 11 and 13) and raising verbs (cf. 12 and 14), as shown in the examples below:

Fornese

11.

A	ì	miei	là	a	fonc-s	diman
EXPL	be.3SG	better	go.INF	to	mashroom-PL	tomorrow

'It's better to go pick up mushrooms tomorrow.'

12.

A	sumiares	ca	a	sepi	da-i	moud-s	par	uda-lu
EXPL	seem.3SG.COND	that	EXPL	be.3SG.SUBJ	of-the	way-PL	for	help.INF-3SG.M.OCL

'There would seem there to be some ways to help him.'

Cilentano

13.

Chiru	pare	rə	ngannarisce	a	bevi
EXPL	seem.3SG	DAT.CL	guzzle.3SG	to	drink.INF

'It seems he likes drinking very much.'

14.

Chiru	è	mala(g)urato	nasce	femmene!
EXPL	be.3SG	unfortunate	born.INF	female.PL

'It is a disgrace to be born women!'

While *a* and *chiru* obligatorily surface in the syntactic environments outlined above, they are optionally found in transitive or unergative clauses featuring a null referential subject. In these contexts, the discourse-pragmatic expletive is in complementary distribution with an overt pronominal or lexical subject. This is shown in Examples (15) to (18) below:

Fornese

15. a.

(A)	n-al	va	mai	four	da-i	peis
EXPL	NEG-3SG.M.SCL	go.3SG	never	out	of-the	foot.PL

'He never leaves.'

b.

Mario	(*a)	n-al	va	mai	four	da-i	peis
Mario	EXPL	NEG-3SG.M.SCL	go.3SG	never	out	of-the	foot.PL

'Mario never leaves.'

16. a. (A) i mangion polenta achi
EXPL 1PL.SCL eat.1PL polenta here
'We eat polenta here.'

b. (*A) nos i mangion polenta achi
EXPL we 1PL.SCL eat.1PL polenta here
'We eat polenta here.'

Cilentano

17. a. (Chiru) penzu avianu sta angora
EXPL think.1SG have.3PL.IMP stay.INF still
'I think that they should have stayed.'

b. Io (*chiru) penzu avianu sta angora
I EXPL think.1SG have.3PL.IMP stay.INF still
'I think that they should have stayed.'

18. a. (Chiru) natàru a puzzu ri Rafeli a lu passatu
EXPL swim.3PL.PST at well of Raffaele to the past
'They swam in Raffaele's well long ago.'

b. (*Chiru) i wagliuni natàru a puzzu ri Rafeli a lu passatu
EXPL the boy.PL swim.3PL.PST at well of Raffaele to the past
'The boys swam in Raffaele's well long ago.'

As shown in Examples (15) to (18), by virtue of being pro-drop languages, in Fornese and Cilentano, the subject can be omitted. The discourse-pragmatic expletives *a* and *chiru* can only optionally surface in the absence of a lexical or pronominal subject. In both languages, the relative order of discourse-pragmatic expletive and overt subject is irrelevant: both linear orders yield an ungrammatical sentence when the two elements co-occur in the same clause. Note that in Examples (15) to (18), agreement on the inflected verb is governed by the null referential subject, as opposed to the surfacing of default third-person singular agreement (cf. 3 to 14).

In Section 3, we will argue that the alternation between those syntactic contexts where the discourse-pragmatic expletive must surface obligatorily (cf. 3–14) and those where it can optionally surface (cf. 15–18) can be constrained in relation to the aboutness of the sentence. More specifically, syntactic contexts featuring an expletive proper in null-subject languages tend to be *thetic* (in the sense of Sasse 1987). In this paper, we use the term *thetic* to refer to a sentence that lacks an XP carrying an aboutness interpretation (i.e., an overt subject or an aboutness/shift topic). In these contexts, Fornese and Cilentano signal zero aboutness through the insertion of *a* and *chiru*, respectively. On the other hand, in transitive and unergative clauses, aboutness is, by default, satisfied by the null or overt subject of the clause (Rizzi and Shlonsky 2006, 2007). In these contexts, the presence of the discourse-pragmatic expletive is optional with a null referential subject. If *chiru* and *a* are present, we argue the effect is a topic shift where a new aboutness/shift topic must be drawn from the propositional content following the discourse-pragmatic expletive.

In her discussion of thetic sentences and expletives, Schaefer (2020, p. 11) argues that a "pragmatically contentful expletive" is a lexical item that is taken from the lexicon to fulfill the discourse function of triggering a thetic interpretation. According to this view, different lexical items with different morpho-syntactic properties can potentially develop into discourse-pragmatic expletives. Indeed, this seems to be the case of *a* and *chiru*. The former is found in nearby closely related Romance varieties as a left-peripheral invariant vocal clitic, whereas the latter may also function as a masculine third-person singular distal demonstrative pronoun in Cilentano. In the next two sections, we will separately look at *a* in Fornese and *chiru* in Cilentano, distinguishing their discourse-pragmatic expletive behavior from the morpho-syntactic behavior of the homophonous elements from which they allegedly developed.

2.2. A in Fornese and across North-Eastern Italian Dialects

An atonic particle *a* has been recorded in several NEIDs as an element of the C-domain. It is argued to mark the following material of the utterance as a new informational broad focus (Benincà 1994, for a in Padovano; Poletto's (2000), discussion of the invariant clitic a; see also Calabrese and Pescarini 2014, for a in the neighboring variety of Forni di Sotto; Casalicchio and Masutti 2015, for a in Campone). NEIDs' *a* is described as a specialized invariant vocal clitic (see Benincà 1994; Poletto 2000; Bernini 2012). In this section, we will show that Fornese *a* is a different element, namely a discourse-pragmatic expletive, which surfaces in the higher portion of the TP-field and behaves like a weak pronominal element (in the sense of Cardinaletti and Starke 1999). The fact that in several NEIDs, *a* has been argued to introduce a thetic sentence suggests that, in principle, the analysis of *a* proposed in this paper may be extensible to other NEIDs; we will, however, leave this point for future research.

All NIDs exhibit full sets or partial sets of obligatory subject clitics (Renzi and Vanelli 1983; Rizzi 1986; Brandi and Cordin 1989). Fornese *a*, given its syntactic distribution, might be regarded as an instance of subject clitic expletive, which obligatorily surfaces in a subset of NIDs in those syntactic environments that require expletive *pro* (see Pescarini 2014, for an overview). In the next few paragraphs, however, we will show that the morpho-syntactic behavior of Fornese *a* is incompatible with the morpho-syntactic behavior of a subject clitic (abbreviated as SCL). Across NIDs, SCLs are phonetically realized pronominal elements that, if required by the grammatical person or the syntactic context, obligatorily accompany finite verbs, as shown in (19) below:

Fornese

19.	a.	As	en	brutas
		3PL.F.SCL	be.3PL	ugly.FPL
		'They are ugly.'		
	b.	*En	brutas	
		be.3PL	ugly.FPL	
		'They are ugly.'		

Due to their unstressed phonological nature, SCLs are also called atonic pronouns to distinguish them from tonic pronominal subjects. Atonic pronouns (or SCLs) are not as free as tonic subject pronouns but have a fixed syntactic position adjacent to the inflected verb (Benincà 1994). In the vast majority of NIDs, tonic and atonic pronouns can co-occur within the same clause. A single inflected verb can hence have two pronominal elements: an obligatory subject clitic and an optional tonic pronoun (Renzi and Vanelli 1983). All NIDs have a set of atonic pronouns, but their number and obligatoriness is subject to cross-dialectal variation. Some NIDs present a complete set of subject clitics (one for each grammatical person), while others have only a partial set that always includes the second-person singular SCL (Benincà 1994). In the literature, SCLs are treated as rich agreement markers between the overt or null subject and the finite verb (Rizzi 1986; Brandi and Cordin 1989; Poletto 2000); nevertheless, evidence from some NIDs shows that, at least in certain varieties, including Paduan, SCLs are *bona fide* resumptive pronouns (see Benincà and Poletto 2004).

As previously mentioned, across NIDs, SCL expletives morphologically mark agreement (or lack of agreement) between a non-referential *pro* and the inflected verb. We argue, however, that this is not the case with Fornese *a*, which instead seems to itself lexicalize a position that is higher than expletive *pro* (Rizzi 1990; Rizzi and Shlonsky 2006, 2007). Fornese *a*, in fact, exhibits a different morpho-syntactic behavior from that of SCLs. For a start, *a* does not undergo subject clitic inversion in root clauses, as shown in (20) and (21):

Fornese

20.	a.	Las	boisas	as	mangia
		the.PL.F	girl.PL.F	SCL.3PL.F	eat.3PL
		'The girls eat.'			
	b.	Mangi-las	las	boisas?	
		eat.3PL-SCL.3PL.F	the.PL.F	girl.PL.F	
		'Do the girls eat?'			
21.	a.	A	ì	tanç	bois
		EXPL	be.3SG	many	boy.PL
		'There are a lot of boys'			
	b.	A	era	tantas	boisas?
		EXPL	be.3SG.PST	many	girl.PL.F
		'Were there many girls?'			

In the existential interrogative in (21b), *a* does not undergo subject clitic inversion; the third-person feminine plural SCL *las* instead obligatorily undergoes inversion with the inflected verb in root interrogatives. Another difference with the syntactic behavior of SCLs is that *a* does not follow the negation but precedes it, as shown in (22):

Fornese

22.	A	na	riva	las	feminas
	EXPL	NEG	arrive.3SG	the.PL.F	woman.PL
	'The wives won't come.'				

23.	Mario	n-al	a	viart	la	puarta
	Mario	NEG-SCL.3SG.M	have.3SG	open.PTCP	the	door
	'Mario did not open the door.'					

SCLs follow the negation in Fornese (cf. 23). This is not the case with *a*, which must instead precede *na*. Furthermore, in negative declarative clauses featuring weather verbs, speakers may also optionally include the SCL expletive *al*, which, if present, follows the negation, as shown in (24):

Fornese

24.	A	na-(l)	maja	mai
	EXPL	NEG-SCL.EXPL.3SG	rain.3SG	never
	'It never rains'			

Across NIDs, it is common for the SCL expletive to have the same form as the referential third-person singular masculine SCL (Renzi and Vanelli 1983; Pescarini 2014). Please note that the subject clitic expletive *al* does not appear in any other morpho-syntactic context in the language. Recall from the previous section that the presence of the SCL expletive *al* with weather verbs is optional; however, the presence of *a* is obligatory in this context. The example in (24) clearly shows that, in Fornese, *a* cannot be considered a SCL expletive, which instead occupies a different syntactic position, following the negation in the same fashion as referential SCLs.

A last piece of evidence comes from coordinated structures, where Fornese *a* does not align with the morpho-syntactic behavior of referential or expletive SCLs. A coordinated clause can only exhibit a single instance of *a*, which cannot be repeated in the second part of the coordinated structure. SCLs, on the other hand, must be obligatorily included in the second part of the coordinated structure (see Rizzi 1986). This is shown in Examples (25) and (26), respectively:

Fornese

25. a. A ì freit e niviê
EXPL be.3SG cold and show.3SG
'It is cold and snows.'

b. *A ì freit e a niviê
EXPL be. 3 SG cold and EXPL show.3SG
'It is cold and snows.'

26. a. Mario al ì rivât e al a mangiât
Mario 3SG.M.SCL be.3SG arrive.PTCP and 3SG.M.SCL have.3SG eat.PTCP
'Mario arrived and ate'

b. *Mario al ì rivât e a mangiât
Mario 3SG.M.SCL be.3SG arrive.PTCP and have.3SG eat.PTCP
'He arrived and ate.'

In this respect, Fornese clitic *a* does not behave like a SCL, as it seems to surface in a position that is higher than that occupied by SCLs.

As previously mentioned, discourse-pragmatic expletive *a* can coexist with a null referential subject (i.e., referential *pro*), but the presence of *a* is incompatible with an overt lexical or pronominal subject. This observation seems to confirm that Fornese *a* sits in a position that is higher than the T° head. At the same time, as we will show in Section 4, *a* appears in a syntactic position lower than left-peripheral focus. Fornese *a* gravitates around the preverbal clitic cluster. Nothing seems to be able to intervene between *a* and the inflectional domain (i.e., TP). *A* cannot be focalized and cannot be used in isolation. In this respect, we can safely conclude that *a* in Fornese does not behave like a proper tonic pronoun either. These facts, together with the morpho-syntactic distribution of *a*, suggest that Fornese *a* is a weak pronominal element (in the sense of Cardinaletti and Starke 1999) that occupies the higher portion of the TP-domain. In Sections 3 and 4, we will corroborate this claim with further evidence. Hence, *a* does not behave like a SCL nor like a tonic pronominal element.

2.3. Chiru in Cilentano and across the Campanian Dialects

The discourse-pragmatic expletive *chiru* in Cilentano developed from the third-person singular masculine distal demonstrative pronoun *chiru*. The pronoun *chiru* has a clear etymology: according to Rohlfs (Rohlfs 2021), Cil. *chiru* < ECCU(M) ILLU(M) Lat., where /r/ comes from the alteration of /ll/ in syntactic protony and is super-extended by analogy to pronominal contexts (see Cerullo 2018). The distal demonstrative pronoun *chiru* is productively used as such in the language, as shown in Examples (27) and (28):

Cilentano

27. A chiru tiempu nisciunu sapìa lègge e scrivi
In that time nobody know.3SG.IMP read.INF and write.INF
'At that time, no one could read or write.'

28. U primu punticjeddu ca è statu fattu?
the first little.bridge that be.3SG be.PTCP make.PTCP
Chiru nun me ricordo.
that NEG me remember.1SG
'The first little bridge that has been built? That I don't remember.'

In (27), *chiru* pre-nominally modifies *tiempu* 'time', whereas in (28), *chiru* is anaphorically bound to the content of the preceding utterance "*U primu punticjeddu ca è statu fattu*" and fronted for discourse-pragmatic reasons (i.e., to assign a contrastive interpretation).

We argue that it is from discourse-pragmatically salient uses of the distal demonstrative pronoun, like (28), that the pronoun *chiru* has developed its discourse-pragmatic expletive value. In other words, the deictic nature of *chiru* contributed towards the establishment of

its function as a marker of zero aboutness. Generally speaking, a demonstrative pronoun is a grammatical word that has a pointing (or deictic) reference (cf. Dixon 2003). The deictic value automatically anchors the pronoun in the universe of speech to its HIC-ET-NUNC, conveying essential pragmatic information as well as a cataphoric or anaphoric relation of identity (Lyons 1977, 1979). In Levinson's (1983, p. 83) terms, this refers to "discourse deixis". The acquired function of the distal demonstrative pronoun *chiru* is hence to contribute "deictically" to the management of the Common Ground (see Stalnaker 1974; Lewis 1979): it signals the absence of an overt or null *aboutness/shift* topic by surfacing as a syntactic–pragmatic placeholder.

Across the Campanian Southern Italian Dialects, Cilentano is not the only variety that exhibits demonstrative pronouns that encode a special pragmatic interpretation. The most exhaustively studied phenomenon is the distal demonstrative *chillo/chello* in double-subject construction in Neapolitan (see Sornicola 1996; Ledgeway 2010; see also Vitolo 2006, for the northern Salerno area).[6] Ledgeway (2010) argues that these structures mark a categorical sentence that serves to establish a new topic. Cilentano exhibits the same type of structure, which is exemplified in (29):

Cilentano

29. Chira la mamma nu bole ca vai ascianne sigarette
DEM.SG.F the mother SG.F not want.3SG that go.3SG ask.INF cigarette.PL
'The mother doesn't want him to go around asking for cigarettes.'

Differently from invariant discourse-pragmatic expletive *chiru*, in these structures, the distal demonstrative pronoun agrees in gender and number with a clause-internal DP, as shown in (29), where *chira* agrees with *la mamma* 'the mother'. The presence of the determiner *la* clearly shows that *chira la mamma* does not form a single DP phrase translatable as 'that mother', where the demonstrative pronoun functions as a pre-nominal modifier. Ledgeway (2010) argues that, in such constructions, the demonstrative pronoun is an element of the C-domain (see also Sornicola 1996), which contextually functions as a "topic-announcing" and "topic-shifting" element. We argue that Ledgeway's (2010) analysis can be extended to Cilentano double-subject constructions like (29).

In this paper, we only investigate invariant *chiru*, which surfaces in the syntactic environments outlined in Section 2.1 above. We will hence not consider those cases in which *chiru* agrees with a clause-internal DP (see Sornicola 1996 and Ledgeway 2010 for further discussion); nevertheless, we want to put forward the idea that double-subject constructions are related to discourse-pragmatic expletive *chiru*, as both strategies ultimately relate to the *aboutness* of the clause. Following Ledgeway (2010), we can rephrase his claim by saying that, in double-subject constructions, the function of the distal demonstrative pronoun is that of signaling that the clause-internal DP with which it agrees must be interpreted as the *aboutness/shift topic* of the clause. We can translate this syntactically into a left-peripheral aboutness discourse feature, which is satisfied through an agree relation with a clause-internal DP, which, in turn, is assigned an aboutness/shift interpretation.[7] In this respect, invariant *chiru*, as a discourse-pragmatic expletive, surfaces in a thetic clause as a last-resort strategy because it lacks a clause-internal element that could be interpreted as an aboutness/shift topic. Demonstratives in double-subject constructions and discourse-pragmatic expletive *chiru* may, therefore, be accounted for through a unitary syntactic analysis; we will nonetheless leave such analysis for future research.

As far as Cilentano invariant *chiru* is concerned, it seems to exhibit a comparable syntactic–pragmatic behavior to Neapolitan distal neuter pronoun *chello* (see Sornicola 1996; Ledgeway 2010). Sornicola (1996) and Ledgeway (2010) argue that, similarly to Cilentano *chiru* and Fornese *a*, *chello* is generally followed by new information (i.e., *broad focus*), and it is incompatible with left-peripheral topicalizations. In this respect, Cilentano diverges from Neapolitan in the use of the third-person singular masculine distal demonstrative pronoun *chiru* as opposed to the neuter form *cheru* as a discourse-pragmatic expletive. Nevertheless, in Cilentano, the neuter form of the demonstrative, *cheru*, is less frequently attested in the

contexts outlined in Section 2.1. It is also important to note that, as a discourse-pragmatic expletive, *cheru* is never possible with meteorological verbs. This suggests that the use of *cheru* as opposed to *chiru* is not interchangeable. We argue that the use of the neuter distal demonstrative pronoun in such contexts brings about a further layer of discourse-pragmatic interpretation, which pertains to the situational context of the utterance from a speaker-related perspective (see Sornicola's (1979), *egocentric reference*; Bartra-Kaufmann 2011; Greco et al. 2017).[8] Let us consider Examples (30) and (31) below:

Cilentano

30. Cheru/#chiru è buono ca veni
DEM.SG.N/EXPL be.3SG good that come.2SG
'The fact is that it is better that you come.'

31. Cheru/#chiru mo aggiu ssuta
DEM.SG.N/EXPL now have.1SG go.out.PTCP.F.SG
'The fact is that I just got back, [I'm sorry].'

In these examples involving *cheru*, the translation is rendered with the expression "the fact is that…". If *chiru* is used instead, this pragmatic layer of interpretation is either lost, or *chiru* is simply interpreted as a regular masculine singular distal pronoun. For example, in (30), if *chiru* were to be used, the interpretation of the sentence would be "that man is good that he comes". Sornicola (1996) argues that Neapolitan expletive-like neuter distal demonstrative *chello* is only allowed in explicative semantic contexts, which can be paraphrased with the expression 'the fact is that…'. In line with Sornicola (1996), we argue that Cilentano *cheru* has the main function of converting the sentence from declarative to explicative, as *chello* does in Neapolitan. Despite the exact discourse-pragmatic nature of *cheru* in Cilentano, which we will not further discuss in this paper, what clearly emerges from this discussion is that Neapolitan seems to lack a true discourse-pragmatic expletive like Fornese *a* and Cilentano *chiru*. In fact, in her comparison of the pragmatic value of the neuter and the masculine distal demostrative pronouns as expletive-like elements, Sornicola (1996) shows that the masculine *chillo* retains some referential traits (i.e., exophorically and endophorically). The discourse-pragmatic expletive use of the third-person singular masculine distal demonstrative pronoun *chiru* hence seems to be an innovation of Cilentano among the Campanian Southern Italian Dialects: *chiru* signals a sentence that lacks a null or overt element that bears an aboutness/shift interpretation. In the next section, we will further explore the notion of aboutness and the interplay of this discourse-pragmatic notion and the manifestation of Fornese *a* and Cilentano *chiru*.

3. Marking Zero Aboutness: A Last-Resort Strategy

In his discussion of the development of Germanic expletive topics, Faarlund (1990) adopts a general notion of *topic*, which can be paraphrased in light of Reinhart's (1981) definition of topic: what the sentence is about. The same is true of the discussion of topic expletive *sitä* in Finnish (Holmberg and Nikanne 2002) and *það* in Icelandic (Rögnvaldsson and Thráinsson 1990). In the last three decades, however, it has been convincingly shown that *topic* is an umbrella term for a non-homogenous class of elements that encode different discourse-pragmatic interpretations and occupy different syntactic positions. There is general agreement in the literature on the existence of at least three (macro-)types of topics: *aboutness/shift topics*, *given* or *familiar topics* and *contrastive topics* (see Frascarelli and Hinterhölzl 2007; Bianchi and Frascarelli 2010). Aboutness/shift topics often mark a shift in conversation; they newly propose or reintroduce a topic in discourse. This type of topic provides a "file card" under which propositional content is stored. In this respect, aboutness/shift topics pertain to *common ground management* (see Krifka 2007; Krifka and Musan 2012): the systematization of the hierarchical organization of the discourse knowledge shared between speaker and hearer. As for given or familiar topics, they instead pertain to *common ground content*: they are contextually given and, therefore, discourse-linked (in the sense of Pesetsky 1987). Given or familiar topics frequently refer to

a pre-established aboutness-shift topic, marking *topic continuity* (Givón 1983; Frascarelli 2017). Lastly, contrastive topics introduce a discourse-related set of alternatives, which are independent of the focus value of the proposition, creating oppositional relations with other topics (Büring 1999). By virtue of encoding different discourse-pragmatic interpretations, we will assume that, at the syntax–pragmatics interface, topical elements are assigned their intended discourse-pragmatic reading through the valuation of specialized topical discourse features: namely, an [Aboutness] topic feature, a [Givenness] topic feature, and a [Contrast] topic feature, respectively. In this section, we will show that the manifestation of discourse-pragmatic expletives *a* and *chiru* is exclusively linked to the satisfaction of a specific type of topical discourse feature, namely [Aboutness]. As previously mentioned, we claim that Fornese and Cilentano must satisfy aboutness structurally by merging an overt or null (i.e., presupposed) aboutness/shift element in the spine of the clause. When no aboutness/shift topic can be elected in discourse (i.e., zero aboutness), *a* and *chiru* are externally merged in the derivation of the clause as a last-resort strategy. In this respect, Fornese *a* and Cilentano *chiru* signal that, in the *common ground*, the following information will not be stored under any specific "file card", and a new aboutness/shift topic must be selected in the propositional content of the sentence.

So far, we have shown that *a* and *chiru* appear in those thetic sentences where a non-null-subject language like English would feature obligatory subject expletives, namely with weather verbs, presentational and existential constructions, impersonal clauses, and in the case of extraposition (see Williams 2001; Biberauer and Roberts 2010; Pescarini 2014). The occurrence of *a* and *chiru* is, however, not limited to those syntactic contexts: we have seen that *a* and *chiru* can also optionally surface in a sentence featuring a null referential subject. In such context, *a* and *chiru* are crucially in complementary distribution with an overt lexical and pronominal subject. This fact neatly shows the link between the aboutness of the clause and the surfacing of the discourse-pragmatic expletive. In null-subject languages, if the subject is overt, in the absence of an established aboutness/shift topic, the lexical or pronominal subject becomes the discourse element that, by default, tells us 'what the sentence is about' (Rizzi and Shlonsky 2006, 2007; Bentley and Cruschina 2018). In these contexts, there is hence no need to resort to the insertion of a discourse-pragmatic expletive, as "default" aboutness is already satisfied by the overt subject. As for those cases where *a* and *chiru* appear with a null referential subject, like in the Fornese example in (32) below, the sentence is interpreted as thetic or, more specifically, lacking established aboutness/shift topic (i.e., zero aboutness):

Fornese

32.

E	ma	"posadas",	ce	dison-as	nos?
And	but	cutlery,	what	say.1PL-SCL.1PL	we
N-i	dison.				
NEG-SCL.1PL	say.1PL				
'As for the word for "cutlery", what do we say? We don't say it.'					
A	tu	dis	diretamentri	i	piron-s
EXPL	2SG-SCL	say.2SG	directly	the	fork-PL
e	la	sidon-s.[9]			
and	the	spoon-PL			
'You can say directly forks and spoons.'					

The presence of the discourse-pragmatic expletive *a* forces a topic shift; in the case of (32), speakers shift away from the established aboutness topic 'cutlery', and a new aboutness topic has to be established from the propositional content that follows the discourse-pragmatic expletive. The optionality of *a* and *chiru* with a null referential subject lies in the fact that not all subjectless clauses receive a zero aboutness interpretation: the aboutness topic may have been previously established in discourse and, hence, be easily retrievable from discourse. In such a case, we argue that [uAboutness] is structurally satisfied by a null topic that is merged in the spine of the clause and anaphorically referential to the

pre-established aboutness/shift topic. Those syntactic environments where an expletive subject would surface in a non-subject language are instead thetic by definition, generally encoding all-new-information (i.e., *broad focus*). No overt or null aboutness/shift topic is available to satisfy the [aboutness] of the sentence; hence, the discourse-pragmatic expletive as a syntactic–pragmatic placeholder surfaces as a last-resort strategy.

The question arises as to why [aboutness] must be satisfied at LF in languages like Fornese and Cilentano. The underlying assumption is that, in line with Erteschik-Shir (1999), the truth value of the propositional content of all clauses must be checked against an aboutness/shift topic. This crucially includes all-new-information sentences in *broad focus* (see Lambrecht 1994), which hence also possess a topic-comment articulation. We argue that those null-subject languages that exhibit discourse-pragmatic expletive elements comparable to *a* and *chiru* belong to a sub-class of languages where this requirement is structurally marked.[10] In the absence of an overt or null (i.e., presupposed) aboutness/shift topic or overt subject that syntactically fills the functional projection responsible for marking "default" aboutness, the discourse-pragmatic expletive obligatorily surfaces to saturate [uAboutness]. The function of these elements is to signal that, with respect to Common Ground management (see Krifka 2007; Krifka and Musan 2012), the information that follows has no "file card" under which to be stored and that an aboutness/shift topic must be established from the propositional content of the following information uttered by the speaker. The surfacing of Fornese *a* and Cilentano *chiru* is hence not incompatible with *any* type of topicalization but only with those overt or null elements that encode an aboutness/shift interpretation. For example, topical frame-setters can co-exist with *a* and *chiru*, as shown in (33) and (34) below:

Fornese

33.	La	setimana	pasada	a	ì	vignût	to	fradi
	the	week	last	EXPL	be.3SG	come.PTCP	your	brother
	Tita	a	ciata-mi					
	Tita	to	find.INF-PRN.1SG					

'Last week, your brother Tita came to visit me.'

Cilentano

34	A	caravanna	chiru	pare	c'anu	mort-e
	At	other.side.of.the.valley	EXPL	seem.3SG	that-have.3PL	die-PTCP.3PL.F
	li	bbacche	ri	Caracca		
	the	cow.PL	of	Caracca.family		

'On the other side of the valley, it seems that the Caraccas' cows died.'

Frame-setting elements provide temporal and/or modal restrictions to the circumstances of evaluation of the proposition expressed by the clause (Haegeman 2000, 2006, 2007; Benincà and Poletto 2004; Poletto 2002). Frascarelli (2017) shows that frames serve a different discourse-pragmatic function than aboutness/shift topics and syntactically occupy a higher position in the left periphery of the clause. As shown in (33) and (34), the spacio-temporal frames *la setimana pasada* 'last week' and *a caravanna* 'on the other side of the valley' can co-occur with *a* and *chiru*, respectively. This is because frames do not strictly encode aboutness. If a frame-setting element is present in a zero aboutness clause, it strictly precedes the discourse-pragmatic expletive.

We will now explore some more evidence in support of our claim that Fornese *a* and Cilentano *chiru* surface as a last-resort strategy to satisfy [uAboutness] in the absence of an overt null aboutness/shift topic. In those syntactic contexts where a non-null-subject language like English features obligatory subject expletives, the presence of *a* and *chiru* seems to be obligatory. These crucially include presentational and existential constructions, which are intrinsically thetic (Sornicola 2010; Bentley et al. 2015). They encode all-new-information (i.e., *broad focus*) and, hence, are felicitous answers to the question "what happened?". Examples (35) to (38) show that if, in an existential or locative construction, the nominal predicate is topicalized, the sentence loses its thetic interpretation: the topicalized

portion of the clause is interpreted as an aboutness/shift topic, and the occurrence of *a* or *chiru* makes the clause ungrammatical:

Fornese

35. a. A ì suiamans tal scansel
EXPL be.3SG towel.PL in-the drawer
'There are towels in the drawer.'

b. *Ì suiamans tal scansel
be.3SG towel.PL in-the drawer
'There are towels in the drawer.'

36. a. I suiamans, i en tal scansel
the towel.PL 3PL.SCL be.3PL in-the drawer
'The towels are in the drawer.'

b. *I suiamans, a ì tal scansel
the towel.PL EXPL be.3SG in-the drawer
'The towels are in the drawer.'

Cilentano

37. a Chiru ave parecchie trote a Calore
EXPL have.3SG many trout at Calore.river
'There are plenty of trout in the Calore river.'

b *Ave parecchie trote a Calore[11]
have.3SG many trout at Calore.river
'There are plenty of trout in the Calore river.'

38. a Parecchie trote, ave a Calore
many trouts, have.3SG at Calore river
'There are plenty of trout in the Calore river.'

b *Parecchie trote chiru ave a Calore[12]
many trouts EXPL have.3SG at Calore.river
'There are plenty of trout in the Calore river.'

Examples (35) and (37) show that, in these contexts, if the pragmatic-expletive is omitted, the sentence is ungrammatical. However, if the nominal predicate is established as the aboutness/shift topic of the clause, *a* and *chiru* cannot surface. Examples (36) and (38) involve an overt aboutness/shift topic, which blocks the surfacing of the discourse-pragmatic expletive; we will now provide two pieces of evidence that show that *a* and *chiru* are also incompatible with an aboutness/shift topic that is presupposed in discourse, and hence null at PF for economy reasons. One such example concerns the topicalization of the partitive argument, which is obligatorily resumed by a reflex of Latin INDE in both Fornese and Cilentano (see INDE cliticization Burzio 1986; Levin and Rappaport-Hovav 1995; Sorace 2000). In such a case, even if the topicalized XP is not phonologically realized (but TP-internally resumed by INDE), the surfacing of *a* and *chiru* is barred. This is shown in Examples (39) to (42), which feature an existential sentence and an unaccusative sentence in both Fornese and Cilentano:

Fornese

39. CONTEXT: Talking about the number of eggs in the fridge.
a. SPEAKER A: I credi ca (*a) nda siepi vuot
1SG.SCL belive.1SG that EXPL PRT.CL be.3SG.SUBJ eight
'I believe there are eight.'
b. SPEAKER B: No, (*a) nd è seis
NEG EXPL PRT.CL be.3SG six
'No, there are six.'

40. CONTEXT: Talking about outsiders (i.e., people from outside the village) in the village.
    a. SPEAKER A: A ì forests in chistu paîs?
       EXPL be.3SG strangers in this village
       'Are there any outsiders in this village?'
    b. SPEAKER B: Si, (*a) nd è.
       Yes EXPL PRT.CL be.3SG
       (*A) nd è rivat-s tanç ist an
       EXPL PRT.CL be.3SG arrive.PTCP-PL many this year
       'Yes, there are. Many have arrived this year.'

Cilentano

41. CONTEXT: Talking about plums on the trees.
    a. SPEAKER A: Creu (*chiru) ngə ne so aulečene fori?
       think.1SG EXPL PF PRT.CL be.3PL plums outside?
       'Are there any in the orchards?'
    b. SPEAKER B: (*Chiru) nun ne ave cchiù!
       EXPL NEG PRT.CL have.3SG more
       'There are no more of them.'

42. CONTEXT: Talking about tourists in the village during summer.
    a. N' ana venuti justu quacchérunu
       PRT.CL have.3PL come.PTCP.3PL only somebody
       'Only some of them came.'

    b. *Chiru nə anu venuti justu quacchérunu
       EXPL PRT.CL have.3PL come.PTCP.3PL only somebody
       'Only some of them came.'

Even if these are syntactic contexts that would require the insertion of an expletive subject in non-null-subject languages, the surfacing of *a* and *chiru* is blocked. This shows that the two discourse-pragmatic expletives do not behave like expletive subjects proper, but their manifestation is constrained by the information structural properties of the clause. The partitive clitics *nd* in Fornese and *ne* in Cilentano are reflexes of Latin INDE, and their use signals that the partitive argument has been topicalized. In Fornese and Cilentano, *nd* and *ne* function as obligatory resumptive pronominal elements. As shown in (39) to (42), in these contexts, the use of *a* and *chiru* is barred: they cannot be realized, as the null topicalized partitive element already satisfies [uAboutness]. In other words, the propositional content of the clause must be stored and interpreted in the Common Ground in light of the presupposed partitive topic.

A further piece of evidence of a null aboutness/shift topic blocking the surfacing of the discourse-pragmatic expletive comes from participial agreement in VS structures in Cilentano. Cerullo (2023) shows that, in Cilentano VS unaccusative structures, if past-participle agreement with the postverbal subject is present, the postverbal subject assumes a discourse-pragmatically salient interpretation; namely, it encodes a topical reading. Depending on the discourse-pragmatic context, the postverbal subject can be cataphorically interpreted as an aboutness/shift topic or anaphorically interpreted as a given/familiar topic (see also De Cia 2022 for an independent analysis of past-participle (in situ) object agreement in Friulian). Consider Example (43) below:

Cilentano

43 a. *(Chiru) a mmuortu Gelsomina
EXPL have.3SG die.PTCP Gelsomina.F.SG
'Gelsomina has died.'

b. (*Chiru) a mort-a Gelsomina
EXPL have.3SG die.PTCP-F.SG Gelsomina.F.SG
'Gelsomina has died.'

According to Cerullo (2023), in (43b), past-participle agreement signals that the postverbal subject *Gelsomina* is discourse-pragmatically salient; in this context, *Gelsomina* is established as the "file card" under which the whole utterance must be interpreted, assuming de facto an aboutness/shift interpretation. The incompatibility with *chiru* can be explained by the fact that, in this case, [uAboutness] is already satisfied by a null left-peripheral aboutness/shift topic in a cataphoric relation to *Gelsomina*.[13] As far as Fornese is concerned, past-participle agreement with the postverbal object or subject of an unaccusative verb is not attested; therefore, this cannot be tested with respect to the surfacing of *a*.

In this section, we have seen that the manifestation of *a* and *chiru* is linked to the satisfaction of aboutness. The discourse-pragmatic expletive lexically marks zero aboutness as a last-resort strategy to saturate the discourse feature [uAboutness]. The investigation of the phenomenon in Fornese and Cilentano suggests that "default" aboutness is not satisfied in the same functional projection in the two languages, but it is subject to parametric variation concerning different functional projections in the syntactic spine of the clause. In the next section, we will show evidence in support of the claim that Cilentano *chiru* surfaces in the higher portion of the C-domain, whereas Fornese *a* surfaces in the higher portion of the extended T-domain.

4. Parametric Choice in the Realization of Zero Aboutness: A Syntactic Account of Fornese *a* and Cilentano *chiru*

Fornese and Cilentano both satisfy [uAboutness] through the lexicalization of the discourse-pragmatic expletives *a* and *chiru* as a last-resort strategy. Nevertheless, we will show that the lexicalization of the discourse-pragmatic expletive does not target the same functional projection in the two languages, but it is subject to parametric variation. *Chiru* seems to be more intimately associated with the C-domain, whereas *a* to the T-domain. Let us consider the examples in (44) and (45) below:

Fornese

44.	Se	a	ì	sot	al	liet?
	What	EXPL	be.3SG	under	the	bed
	'What is there under the bed?'					

Cilentano

45.	a.	*Che	chiru	ng	è	sotta	au	ljettu?
		What	EXPL	PF	be.3SG	under	the	bed
		'What is there under the bed?'						
	b.	Chiru	che	ng	è	sotta	au	ljettu?
		EXPL	what	PF	be.3SG	under	the	bed
		'What is there under the bed?'						

The example in (44) shows that, in an existential *wh*-interrogative clause, Fornese *a* appears in a position lower than the landing position of the *wh*-element *se* 'what'. Abiding by a model of the split C-domain across NIDs that lacks a topic position lower than focus (Benincà and Poletto 2004), it is safe to assume that *a* sits in a functional projection lower than FocP (i.e., the landing site of wh-items, see Rizzi 1997) but above T where the copula sits (see Manzini and Savoia 2005; and Roberts 2010, on the generalized V-to-T movement in Northern Italian Dialects). In Cilentano, on the other hand, the same constituent order yields the ungrammatical sentence in (45a). In the root wh-interrogative in (45), *chiru* cannot appear in a position lower than the wh-item *che* 'what'. (45b) shows that *chiru* must surface in a position that is higher than the landing site of the *wh*-element, namely higher than FocusP.[14] It is, hence, safe to assume that the discourse-pragmatic expletive *chiru* is an element of the C-domain. More specifically, we argue that *chiru* lexicalizes the left-peripheral projection of the *topic field* (in the sense of Benincà and Poletto 2004), which hosts topics bearing an aboutness/shift interpretation, namely ShiftP (see Frascarelli and Hinterhölzl 2007; Bianchi and Frascarelli 2010). This is shown by the distribution of *chiru* in

(46), (47), and (48) below. Please note that the equivalent Fornese sentences in (49), (50), and (51) are additionally provided for comparison:

Cilentano

46. M'addummanu chiru si vène Maria musera
RFLX-ask.1SG EXPL if come.3SG Maria tonight
'I wonder whether Maria will come tonight.'

47. Chiru pecché ana mortu tutti i bbacchi?
EXPL why have.3PL die.PTCP all the cows?
'Why have all the cows died?'

48. Penzo ca chiru nu bbene Maria musera
think.1SG COMP EXPL not come.3SG Maria tonight
'I think Maria will not come tonight.'

Fornese

49. N-i sai se a í pursiei t-al stali
NEG-1SG.SCL know.1SG if EXPL be.3SG pigs in-the barn
'I don't know if there are pigs in the barn.'

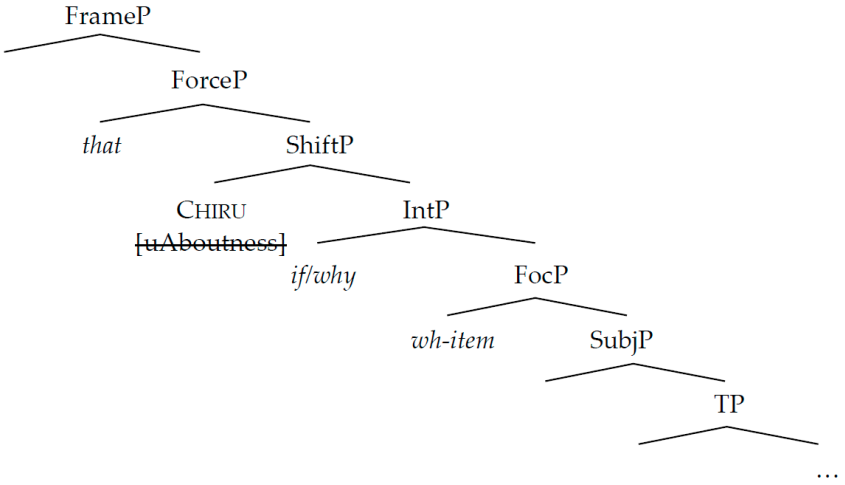
50. Parsè ca a ì muart las bestias?
Why that EXPL be.3SG die.PTCP the cattle
'Why did the cattle die?'

51. I credi ca a era calchidun
1SG.SCL believe.1SG COMP EXPL be.3SG.PSTsomeone
ca ti spietava di four
REL 2SG.PRN wait.3SG.IMP of outside
'I believe there was someone waiting for you outside.'

The examples in (46) and (47) show that *chiru* appears in a position higher than *si* 'if' and *pecchè* 'why'. According to Rizzi (2013, 2018), these two elements lexicalize InterrogativeP (IntP) in the C-domain, which is a functional projection sandwiched by different topical functional projections within the topic field (à la Benincà and Poletto 2004). *Chiru*, hence, surfaces in the higher portion of the topic layer in the left periphery. This can be further appreciated by the fact that *chiru* is realized in a position lower than the complementizer *che* in (5), which lexicalizes ForceP (see Rizzi 1997).[15] This position is indeed compatible with ShiftP within the topic field (Frascarelli and Hinterhölzl 2007; Bianchi and Frascarelli 2010). In Cilentano, zero aboutness is hence marked (i.e., [uAboutness] satisfied) in the canonical C-domain position, which hosts aboutness/shift topics. As for Fornese *a*, Examples (44) and (49) to (51) show that its lexicalization is featured in a lower functional projection above T. We argue that this position is SubjP, the canonical position of overt subjects (Rizzi and Shlonsky 2006, 2007), which also by default can satisfy [uAboutness] in the absence of an established aboutness/shift topic. This is a node between the C-domain and the T-domain, which is, for instance, compatible with the position of *a* in Fornese. Rizzi and Shlonsky (2007) argue that an expletive proper resolves the tension between the formal syntactic requirement of the clause and discourse conditions. For instance, if the thematic structure of a verb requires a presentational structure, in which, by nature, the event described is not "about" something, the expletive subject signals that the clause has to be interpreted presentationally, and no argument is expressed in aboutness position. The behavior of the discourse-pragmatic expletive *a* in Fornese can be analogously seen as satisfying a syntactic requirement whereby, if no overt or null aboutness XP is present, zero aboutness must be phonologically realized. It is important to note that Casalicchio and Masutti (2015) independently show that, in the nearby variety of Campone, *a* lexicalizes Subject° head by virtue of bearing [+ third person] feature. In Fornese, *a* is found in complementary distribution with overt lexical and pronominal subjects, which suggests that, in this variety, *a* sits in the specifier position of SubjectP.[16]

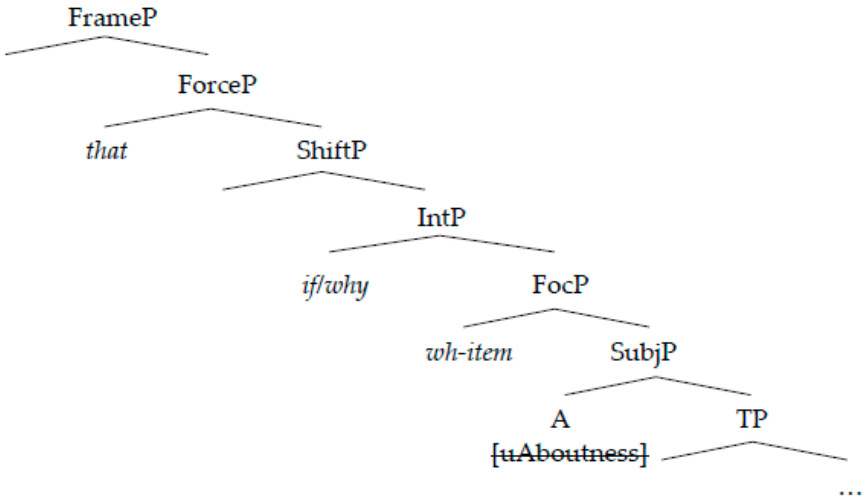
Despite their different syntactic positions, both *a* and *chiru* mark zero aboutness. The lexicalization of *a* and *chiru* is linked to the satisfaction of a default [uAboutness] feature. The fact that, in Fornese and Cilentano, these elements occupy different syntactic positions is evidence that the discourse feature can be satisfied in different functional projections in the clausal spine, namely in the high-TP layer and the C-domain. This is schematically represented in the simplified arboreal representations in (52) and (53) below, which summarize our discussion so far. Note that (52) captures the syntactic behavior of Cilentano *chiru*, whereas (53) refers to Fornese *a*:

(52)



As far as the syntactic analysis of the two discourse-pragmatic expletives is concerned, we argue that Fornese *a* and Cilentano *chiru* are externally merged in the specifier position of SubjP and ShiftP, respectively, as a last-resort strategy to satisfy [uAboutness]. We claim that, in Fornese and Cilentano, the satisfaction of [uAboutness] is a structural requirement at LF. [uAboutness] is an uninterpretable and unvalued feature that must be deleted to prevent the syntactic derivation from crashing (Chomsky 2001, 2004).[17] In the absence of an overt or null XP in the spine of the clause that values and deletes the discourse feature, a discourse-pragmatic expletive is externally merged to rescue the derivation.

(53)



The postulation of such parametric variation, however, does not mean that, in Fornese, aboutness/shift topics by default occupy SubjP and that Cilentano cannot project SubjP to host lexical or pronominal subjects. Both functional projections ShiftP and SubjP are

available in the two languages to host aboutness/shift topics and overt subjects, respectively. In fact, we claim that both elements can satisfy [uAboutness] in the two languages. We want to limit our claim to the existence of a parametric choice in the syntactic locus where zero-aboutness is lexicalized (i.e., the discourse-pragmatic expletive is merged) within the syntactic spine of the clause. Assuming Chomsky's (2001, 2004) probe-goal model, by virtue of being uninterpretable, [uAboutness] can be either the probe or the goal of the syntactic operation *Agree*. This has important structural consequences for the saturation of the discourse feature. We claim that, in Fornese and Cilentano, [uAboutness] can be satisfied by either merging an overt or null aboutness/shift topic in ShiftP or an overt subject in SubjP: what is language-specific—and subject to parametric choice—is the functional head to which [uAboutness] is associated, namely Shift° in Cilentano and Subj° in Fornese. More concretely, we assume three possible ways in which Fornese and Cilentano can satisfy the structural requirement on [uAboutness]. In Cilentano, this involves either (a) externally merging an aboutness/shift topic in ShiftP; or (b) merging a lexical or pronominal subject in SubjP, whose default aboutness interpretation, namely [iAboutness], values and deletes [uAboutness] that, in Cilentano, is part of the lexical specification of the probing Shift° head. If both strategies are unavailable (i.e., in the case of zero aboutness), c) Cilentano lexicalizes the discourse-pragmatic expletive *chiru* in SpecShiftP, which satisfies [uAboutness] on Shift° head. Similarly, in Fornese, [uAboutness] on Subj° head can be satisfied by (a) either merging a lexical or pronominal overt subject in SubjP or (b) externally merging an aboutness/shift topic in ShiftP. In the latter case, if ShiftP is projected, it is also endowed with an unvalued [uAboutness] discourse feature, which probes down to find and agree with the active goal [uAboutness] in Subj°. Via Agree, the topical XP merged in the specifier position of ShiftP also satisfies [uAboutness] in Subj°. In case neither strategy is available (i.e., a or b), (c) Fornese satisfies [uAboutness] in Subj° by externally merging the discourse-pragmatic expletive *a* in SpecSubjP as a last-resort strategy.

Finally, it is important to note that our analysis does not preclude the possibility that the sentence simultaneously hosts an aboutness/shift topic and an overt lexical or pronominal subject. In this respect, our *Agree* account may have interesting implications for the analysis of double-subject constructions in Cilentano (see Section 2.3 above), where an inflected form of the distal demonstrative pronoun *chiru* agrees with a TP-internal DP, which, in turn, assumes a salient discourse-pragmatic interpretation; nevertheless, we will not explore this further, but leave it for future research. The case of *a* and *chiru* in Fornese and Cilentano suggests that, in those null-subject languages where zero aboutness must be overtly marked, there exists a parametric choice with respect to the functional projection, which by default lexicalizes [uAboutness], namely SubjP in Fornese and ShiftP in Cilentano.

5. Conclusions

To conclude, we have shown that *a* and *chiru* are not bona fide expletive subjects but discourse-pragmatic expletives that signal the absence of an aboutness/shift referent in the utterance (i.e., zero aboutness). In the absence of an established overt or null aboutness/shift topic in the common ground, they lexicalize a syntactic position in the syntactic spine of the clause. In other words, they signal that there is no "file card" under which the propositional content of the utterance can be stored, and hence, a new aboutness/shift topic must be selected from the content of the utterance itself or the following utterances. We have also shown that the morpho-syntactic behavior of *a* and *chiru* as discourse-pragmatic expletives differs from that of a subject clitic in Fornese and a referential distal demonstrative pronoun in Cilentano.

In this paper, we claim the existence of a sub-class of null-subject languages where [uAboutness] as a discourse feature must be structurally satisfied by merging an overt or null topic in the syntactic spine of the clause. In the absence of an element that encodes aboutness, a discourse-pragmatic expletive is externally merged as a last-resort strategy. We have argued that the satisfaction of the uninterpretable discourse feature [uAboutness] is an LF requirement, which, cross-linguistically, is subject to a parametric choice. We

show that, in Fornese, "default" [aboutness] is satisfied in SubjP, which is the canonical syntactic position for overt subjects within a cartographic approach. In Cilentano, on the other hand, [aboutness] is satisfied in a higher position within the C-domain, namely ShiftP, the canonical syntactic position that hosts overt aboutness/shift topics. In this respect, we expect to find other null-subject languages exhibiting expletive-like elements that abide by the same parametric choice. For example, the distribution of topic expletive *sitä* in Finnish (Holmberg and Nikanne 2002) seems to closely resemble the morpho-syntactic distribution of Fornese *a*; on the other hand, the distribution of Catalan expletive-like *ell* and Spanish *ello* (see Bartra-Kaufmann 2011) seems to follow more closely that of Cilentano *chiru*. This is captured by the parametric hierarchy that we propose below:

54. THE ZERO-ABOUTNESS HYPOTHESIS
    Zero Aboutness is defined as the absence of an aboutness referent in an utterance.
    i. Zero Aboutness {may/may not} be marked overtly.
    ii. If marked overtly, it is syntactically realized in either ShiftP or SubjP.

The analysis put forward in this paper contributes towards refining our understanding of what expletives are, as well as having interesting repercussions on the traditional view of expletives as purely structural placeholders that are semantically vacuous. Although Fornese *a* and Cilentano *chiru* surface to saturate a formal syntactic feature like subject expletives proper do, their presence or absence does have an interpretative effect on the utterance. In this respect, expletive elements can be seen as functional elements whose interpretative effects are linked to the syntactic feature they lexicalize. Finally, this paper has provided a fresh perspective on the types of "speaker-related meanings" that are available in those null-subject languages that exhibit expletive-like elements. The surfacing of Fornese *a* and Cilentano *chiru* is intrinsically linked with one type of topicality, namely aboutness. This can be, by default, satisfied in two different functional projections in the languages under investigation (i.e., ShiftP and SubjP). For future research, it would be important to investigate whether there are any further structural consequences that are linked to this parametric choice.

Author Contributions: Conceptualization, S.D.C. and M.C.; formal analysis, S.D.C.; investigation, S.D.C. and M.C.; resources, S.D.C. and M.C.; data curation, S.D.C. and M.C.; writing—original draft preparation, S.D.C.; writing—review and editing, S.D.C. and M.C.; visualization, S.D.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Recordings are available from the authors upon request.

Conflicts of Interest: The authors declare no conflict of interest.

Notes

1 These varieties have been traditionally called *Italian Dialects* in the literature. However, this label is misleading, as these Italo-Romance varieties are not dialects of Italian. They are instead independent and autonomous continuations of the vulgar Latin spoken in the Italian peninsula around the 11th century, and hence sister languages of Italian (see Maiden and Parry 1997).

2 With reference to Chomsky's (1981, 1995) model of grammar, it is important to note that the satisfaction of [uAboutness] is not a PF requirement, but, rather, it concerns LF and its interaction with syntax. In fact, the aboutness of the clause can be covertly satisfied by a null presupposed topic that is easily retrievable in discourse.

3 While the absence of *a* in these contexts leads to clear-cut ungrammaticality judgments in Fornese, in Cilentano, speakers seem to be more flexible; nevertheless, in naturally occurring speech, *chiru* in these contexts is virtually always present.

4 Note, however, that in some varieties of Cilentano in the Calore valley, the presence of the proform *ngi* in the existential-locative construction blocks a *broad focus* reading on the existential construction, making the use of the discourse-pragmatic expletive *chiru* ungrammatical. In these varieties, we claim that the existential-locative proform is desemanticized but not fully grammaticalized,

acting de facto as a resumptive pronominal element for a topicalized locative XP. As we will argue in Section 3, the same behavior is attested in both Fornese and Cilentano with INDE-cliticization with a silent partitive topicalized: at LF, the topicalized null XP satisfies [uAboutness], blocking the lexicalization of the discourse-pragmatic expletive.

5 Please note that *cheru*, namely the neuter form of the third-person singular masculine distal demonstrative, cannot be used with meteorological verbs in Cilentano. In Section 2.3, we will further discuss the difference between the masculine and the neuter forms when functioning as discourse-pragmatic expletives.

6 We will use the orthographic representation *chillo*/*chello*. These elements are pronounced as [kil:ə]/[kel:ə]. The pronunciation involves the reduction of the final atonic vowel to schwa [ə]. The only system of disambiguation between the masculine and the neuter form is through metaphony of the tonic vowel. The reduction of the final atonic vowel to schwa is a widespread phenomenon in northern Campanian varieties (see De Blasi 2006).

7 This could be achieved through *Agree* in Chomsky's (2001) probe-goal model. Nevertheless, in this paper, we will not pursue the analysis of this phenomenon any further and direct the reader to Sornicola (1996), Vitolo (2006), and Ledgeway (2010).

8 The alternation between *chiru* and *cheru* may be rooted in diachrony with respect to *esse* vs. *habere* auxiliary selection (see also Cennamo and Cerullo 2021).

9 Please note that *tu* here is a subject clitic (i.e., an agreement marker); it is not a tonic subject pronoun (see discussion in Section 2.2).

10 See, for instance, *sitä* in Finnish (Holmberg and Nikanne 2002), which exhibits an identical morpho-syntactic behavior as *a* in Fornese. Catalan expletive-like *ell* and Spanish *ello* (see Bartra-Kaufmann 2011) instead seem to more closely resemble the syntactic distribution of Cilentano *chiru*.

11 Note again that in some varieties of Cilentano in the Calore valley, the presence of the proform *ngi* in the existential-locative construction blocks a *broad focus* reading on the existential construction, making the use of the discourse-pragmatic expletive *chiru* ungrammatical in such context. In these varieties, we claim that the existential-locative proform is desemanticized but not fully grammaticalized, acting de facto as a resumptive pronominal element for a topicalized locative XP, which encodes an aboutness/shift interpretation. This is shown in (i) below:

(i)	(*Chiru)	ngi	ave	parecchie	trote	a	Calore
	EXPL	PF	have.3SG	many	trout	at	Calore.river
	'There are many trout in the river.'						

12 It is important to note that, here, *chiru* is not a distal demonstrative pronoun, which anaphorically refers to the wealth of trout in the river or, cataphorically, to the river Calore itself.

13 In case *Gelsomina* in (43b) were to bear a *givenness* interpretation (as opposed to an *aboutness/shift* interpretation), the surfacing of *chiru* would only be banned in the case where the speaker wants to mark *topic continuity* (Givón 1983), meaning that *Gelsomina* remains the established aboutness/shift topic in discourse. If the speaker instead wants to shift topic to an XP other than "given" *Gelsomina*, then *chiru* would be able to surface (see De Cia 2022; De Cia et al. 2022; Cerullo 2023, for further discussion on the interaction between discourse features and past-participial agreement).

14 Note that the ability of *a* and *chiru* to co-occur with a wh-item (i.e., an XP in narrow focus; see Lambrecht 1994) makes us tear apart the notion of *zero aboutness* from that of a *thetic* sentence. Zero aboutness simply means that the utterance lacks an overt or null aboutness/shift referent, whereas a thetic sentence, in the strict sense, would not allow the presence of an element in narrow focus, but the whole sentence would have to be in broad focus. *A* and *chiru* hence mark zero aboutness, where the sentence does not necessarily have to be thetic (i.e., all new informational focus).

15 It is important to note that in (46), (48), (49), and (51), the discourse-pragmatic expletive occurs in an embedded clause. This leads to the question of whether all embedded clauses are endowed with the same discourse-related requirements as root clauses. More specifically, in our case, this concerns the satisfaction of [uAboutness]. Our data show that this is indeed the case with embedded clauses introduced by *bridge verbs* (e.g., *say, think, believe*), which present a more articulated left-peripheral space, which exhibits syntactic and discourse-pragmatic phenomena proper of root clauses (Vikner 1995; Poletto 2000; Ledgeway 2008; González i Planas 2014; a.o.). Embedded clauses with a reduced or defective CP layer (e.g., introduced by factive verbs) may not be subject to the same requirement with respect to the satisfaction of aboutness. This was not systematically investigated in this study.

16 The analysis of *a* in Fornese that we propose in this paper is compatible with the claim that, in thetic sentences, a silent subject of predication (SoP) is present in the syntactic spine of the clause, as outlined by Bentley and Cruschina (2018). Schaefer (2020) argues that the SoP syntactically behaves like a null expletive-like element. In this respect, Fornese *a* could be seen as the lexicalization of the SoP. Nevertheless, we claim that this does not apply to Cilentano *chiru*, as we have shown that *chiru* surfaces in a syntactic position that is higher than SubjP, which is the SoP syntactic position postulated by Bentley and Cruschina (2018).

17 See also Ojea's (2017) *core intentional features* for the derivation of the subject in Spanish thetic sentences.

References

Alexiadou, Artemis, and Elena Anagnostopoulou. 1998. Parametrizing AGR: Word Order, V-Movement and EPP-Checking. *Natural Language & Linguistic Theory* 16: 491–539.

Bartra-Kaufmann, Anna. 2011. Recycled neuter expletive pronouns and the nature of left periphery: *Ell* and *relatives* revisited. *Catalan Journal of Linguistics* 10: 185–219. [CrossRef]
Benincà, Paola. 1994. *La variazione sintattica. Studi di dialettologia romanza*. Bologna: Il Mulino.
Benincà, Paola, and Cecilia Poletto. 2004. Topic, focus and V2: Defining the CP sublayers. In *The Structure of IP and CP. The Cartography of Syntactic Structures*. Edited by Luigi Rizzi. Oxford and New York: Oxford University Press, vol. 2, pp. 52–75.
Benincà, Paola, and Laura Vanelli. 2016. Friulian. In *The Oxford Guide to the Romance Languages*. Edited by Adam Ledgeway and Martin Maiden. Oxford: Oxford University Press, pp. 139–53.
Bentley, Delia, and Silvio Cruschina. 2018. The silent argument of broad focus: Typology and predictions. *Glossa: A Journal of General Linguistics* 3: 1–37. [CrossRef]
Bentley, Delia, Francesco M. Ciconte, and Silvio Cruschina. 2015. *Existentials and Locatives in Romance Dialects of Italy*. Oxford: Oxford University Press.
Bernini, Giuliano. 2012. Il clitico *a* nell'italo-romanzo settentrionale: Osservazioni metodologiche. In *Per Roberto Gusmani. Linguistica storica e teorica Studi in ricordo*. Edited by Vincenzo Orioles. Udine: Forum, pp. 269–82.
Bianchi, Valentina, and Mara Frascarelli. 2010. Is Topic a Root Phenomenon? *Iberia: An International Journal of Theoretical Linguistics* 2: 43–88.
Biberauer, Theresa, and Ian Roberts. 2010. Subjects, Tense and Verb-movement. In *Parametric Variation: Null Subjects in Minimalist Theory*. Edited by Theresa Biberauer, Anders Holmberg, Ian Roberts and Michelle Sheehan. Cambridge: Cambridge University Press, pp. 263–303.
Bolinger, Dwight. 1977. *Meaning and Form*. London: Longman.
Brandi, Luciana, and Patrizia Cordin. 1989. Two Italian dialects and the Null Subject Parameter. In *The Null Subject Parameter*. Studies in Natural Language and Linguistic Theory 15. Edited by Osvaldo A. Jaeggli and Kenneth J. Safir. Dordrecht: Kluwer, pp. 111–42. [CrossRef]
Büring, Daniel. 1999. Topic. In *Focus: Linguistic, Cognitive, and Computational Perspectives*. Edited by Peter Bosch and Rob van der Sandt. Cambridge: Cambridge University Press, pp. 142–65.
Burzio, Luigi. 1986. *Italian Syntax: A Government-Binding Approach*. Dordrecht: Kluwer.
Calabrese, Andrea, and Diego Pescarini. 2014. Clitic metathesis in the Friulian dialect of Forni di Sotto. *Probus* 26: 275–308. [CrossRef]
Cardinaletti, Anna, and Lori Repetti. 2010. Proclitic vs. enclitic pronouns in northern Italian dialects and the null-subject parameter. In *Syntactic Variation. The Dialects of Italy*. Edited by Roberta D'Alessandro, Adam Ledgeway and Ian Roberts. Cambridge: Cambridge University Press, pp. 119–34.
Cardinaletti, Anna, and Michal Starke. 1999. The typology of structural deficiency. A case study of the three classes of pronouns. In *Clitics in the Languages of Europe*. Edited by Henk van Riemsdijk. Berlin: Mouton De Gruyter, pp. 145–233.
Carrilho, Ernestina. 2008. Beyond doubling: Overt expletives in European Portuguese dialects. In *Microvariation in Syntactic Doubling*. Edited by Sjef Barbiers, Olaf Koeneman, Marika Lekakou and Margreet van der Ham. Bingley: Emerald, pp. 301–23.
Casalicchio, Jan, and Vania Masutti. 2015. A syntactic analysis of the subject clitic *a* in the Friulian variety of Campone. *Isogloss. Special Issues in Italo-Romance Morphosyntax*, 103–32. [CrossRef]
Cennamo, Michela, and Mariangela Cerullo. 2021. I costrutti esistenziali nelle varietà del Cilento interno. In *Miscellanea di studi in onore di Diego Poli*. Edited by Francesca Chiusaroli. Roma: Il Calamo, pp. 263–80. ISBN 9788898640645.
Cerullo, Mariangela. 2018. La variazione fonetica degli esiti di -LL- latina. Processi di retroflessione (e deretroflessione) e rotacizzazione in alcuni punti della Val Calore Lucano. Master's thesis, University of Naples "Federico II", Naples, Italy.
Cerullo, Mariangela. 2023. Struttura e uso degli ausiliari perfettivi e sistemi di accordo participiale nel cilento. Il continuum areale tra conservatività e innovazione, una prospettiva sociolinguistica. Ph.D. thesis, University of Naples "Federico II", Naples, Italy.
Chelliah, Shobhana, and Willem de Reuse. 2011. *Handbook of Descriptive Linguistic Fieldwork*. Berlin: Springer.
Chomsky, Noam. 1981. *Lectures on Government and Binding*. Dordrecht: Foris.
Chomsky, Noam. 1995. *The Minimalist Program*. Cambridge: MIT Press.
Chomsky, Noam. 2001. Derivation by phase. In *Ken Hale: A Life in Language*. Edited by Michael J. Kenstowicz. Cambridge: MIT Press, pp. 1–52.
Chomsky, Noam. 2004. Beyond explanatory adequacy. In *Structures and Beyond. The Cartography of Syntactic Structures*. Edited by Adriana Belletti. Oxford: Oxford University Press, pp. 104–31.
Cinque, Guglielmo. 1999. *Adverbs and Functional Heads*. Oxford and New York: Oxford University Press.
Cruschina, Silvio. 2012. *Discourse-Related Features and Functional Projections*. Oxford and New York: Oxford University Press.
De Blasi, Nicola. 2006. *Profilo linguistico della Campania*. Bari: Laterza.
De Cia, Simone. 2022. Topicality and Object-Past Participle Agreement in Friulian. In *Angles of Object Agreement*. Edited by Andrew Nevins, Anita Peti-Stantić, Mark de Vos and Jana Willer-Gold. Oxford: Oxford University Press, pp. 295–326. ISBN 9780192897749.
De Cia, Simone, Francesco M. Ciconte, and Delia Bentley. 2022. Participle Agreement with Topical In-situ Objects: The Case of Old Italian. *Archivio Glottologico Italiano* 107: 37–81.
Dixon, Robert M. W. 2003. Demonstratives: A cross-linguistic typology. *Studies in Language* 27: 61–112. [CrossRef]
Erteschik-Shir, Nomi. 1999. Focus structure and scope. In *The Grammar of Focus*. Edited by Georges Rebuschi and Laurice Tuller. Amsterdam: John Benjamins, pp. 119–50.
Faarlund, Jan T. 1990. *Syntactic Change: Toward a Theory of Historical Syntax*. Berlin and New York: De Gruyter Mouton.

Frascarelli, Mara. 2017. Dislocations and framings. In *Manual of Romance Morphosyntax and Syntax, Manuals of Romance Linguistics 17*. Edited by Andreas Dufter and Elisabeth Stark. Berlin and New York: Mouton de Gruyter, pp. 472–501.
Frascarelli, Mara, and Roland Hinterhölzl. 2007. Types of Topics in German and Italian. In *On Information Structure, Meaning and Form*. Edited by Kerstin Schwabe and Susanne Winkler. Amsterdam: John Benjamins, pp. 87–116.
Givón, Talmy. 1983. Topic continuity in discourse: An introduction. In *Topic Continuity in Discourse: A Quantitative cross Language Study*. Edited by Talmy Givón. Amsterdam and Philadelphia: John Benjamins, pp. 5–41.
González i Planas, Francesc. 2014. On Quotative Recomplementation: Between Pragmatics and Morphosyntax. *Lingua* 146: 39–74. [CrossRef]
Greco, Ciro, Liliane Haegeman, and Trang Phan. 2017. Expletive and speaker-related meaning. In *Order and Structure in Syntax II: Subjecthood and Argument Structure*. Edited by Michelle Sheehan and Laura R. Bailey. Berlin: Language Science Press, pp. 69–93.
Haegeman, Liliane. 2000. Inversion, non-adjacent inversion and adjuncts in CP. *Transactions of the Philological Society* 98: 121–60. [CrossRef]
Haegeman, Liliane. 2006. Argument fronting in English, Romance CLLD and the left periphery. In *Negation, Tense and Clausal Architecture: Cross-Linguistic Investigations*. Edited by Raffaella Zanuttini, Héctor Campos, Elena Herburger and Paul Portner. Georgetown: Georgetown University Press, pp. 27–52.
Haegeman, Liliane. 2007. Operator movement and topicalisation in adverbial clauses. *Folia Linguistica* 41: 279–325. [CrossRef]
Himmelmann, Nikolaus. 1998. Documentary and descriptive linguistics. *Linguistics* 36: 161–95. [CrossRef]
Himmelmann, Nikolaus. 2006. Language documentation: What is it and what is it good for? In *Essentials of Language Documentation*. Edited by Jost Gippert, Nikolaus Himmelmann and Ulrike Mosel. Berlin: Walter de Gruyter.
Hinzelin, Marc-Olivier, and Georg A. Kaiser. 2007. El pronombre ello en el léxico del español dominicano. In *Language Contact and Language Change in the Caribbean and Beyond/Lenguas en contacto y cambio lingüistico en el Caribe y más allá*. Edited by Wiltrud Mihatsch and Monika Sokol. Frankfurt am Main: Lang, pp. 171–88.
Holmberg, Anders, and Urpo Nikanne. 2002. Expletives, Subjects, and Topics in Finnish. In *Subjects, Expletives, and the EPP*. Edited by Peter Svenonius. New York: Oxford University Press, pp. 71–106.
Kaiser, Georg, and Eva-Maria Remberger. 2009. *Proceedings of the Workshop "Null Subjects, Expletives, and Locatives in Romance"*. Konstanz: Fachbereich Sprachwissenschaft der Universität Konstanz.
Krifka, Manfred. 2007. Basic Notions of Information Structure. In *Interdisciplinary Studies of Information Structure: The Notions of Information Structure*. Edited by Caroline Fery, Gisbert Fanselow and Manfred Krifka. Potsdam: Universitätsverlag Potsdam, pp. 13–56.
Krifka, Manfred, and Renate Musan. 2012. Information Structure: Overview and Linguistic Issues. In *The Expression of Information Structure*. Edited by Manfred Krifka and Renate Musan. Berlin and Boston: De Gruyter Mouton, pp. 1–44.
La Fauci, Nunzio. 1988. *Oggetti e Soggetti nella Formazione della Morfosintassi Romanza*. Pisa: Giardini.
Lambrecht, Knud. 1994. *Information Structure and Sentence Form: Topic, Focus, and The Mental Representation of Discourse Referents*. Cambridge: Cambridge University Press.
Lambrecht, Knud. 2000. When subjects behave like objects: An analysis of the merging of S and O in sentence-focus constructions across languages. *Studies in Language* 24: 611–82. [CrossRef]
Lambrecht, Knud, and Maria Polinsky. 1997. Typological variation in sentence-focus constructions. In *Thirty-Third Annual Meeting of the Chicago Linguistic Society. Papers from the Panels*. Edited by Kora Singer, Randall Eggert and Gregory Anderson. Chicago: Chicago University Press, pp. 189–206.
Ledgeway, Adam. 2008. Satisfying V2: Sì Clauses in Old Neapolitan. *Journal of Linguistics* 44: 437–40. [CrossRef]
Ledgeway, Adam. 2010. Subject licensing in CP: The Neapolitan double-subject Construction. In *Mapping the Left Periphery*. Edited by Paola Benincà and Nicola Munaro. Oxford: Oxford University Press, pp. 257–96.
Levin, Beth. 2015. *Event Encoding in a Crosslinguistic Perspective IV: Talking about the Weather*. LSA Institute Course 319. Chicago: University of Chicago.
Levin, Beth, and Malka Rappaport-Hovav. 1995. *Unaccusativity: At the Syntax-Lexical Semantics Interface*. Cambridge: The MIT Press.
Levinson, Stephen C. 1983. *Pragmatics*. Cambridge: Cambridge University Press.
Lewis, David. 1979. Scorekeeping in a language game. *Journal of Philosophical Logic* 8: 339–59. [CrossRef]
Lyons, John. 1977. *Semantics*. Cambridge: Cambridge University Press.
Lyons, John. 1979. Deixis and Anaphora. In *The Development of Conversation and Discourse*. Edited by Terry Myers. Edinburgh: Edinburgh University Press, pp. 88–103.
Maiden, Martin, and Mair Parry. 1997. *The Dialects of Italy*. London: Routledge.
Manzini, Maria Rita, and Leonardo Savoia. 2005. *I Dialetti Italiani e Romanci. Morfosintassi Generativa*. Alessandria: Edizioni dell'Orso.
Milroy, Lesley, and Matthew Gordon. 2003. *Sociolinguistics: Method and Interpretation*. London and Melbourne: Blackwell.
Nicolis, Marco. 2008. The Null Subject Parameter and correlating properties: The case of Creole languages. In *The Limits of Syntactic Variation*. Edited by Theresa Biberauer. Amsterdam: Benjamins, pp. 271–94.
Ojea, Ana. 2017. Core intentional features in the syntactic computation: Deriving the position of the subject in Spanish. *Lingua* 195: 72–91. [CrossRef]
Pellegrini, Giovan Battista. 1979. I dialetti ladino-cadorini. *Archivio per l'Alto Adige* 72: 245–65.
Pescarini, Diego. 2014. La distribuzione dei clitici soggetto espletivi: Tipologia e prospettive parametriche. *L'italia Dialettale* 75: 229–46.

Pesetsky, David. 1987. Wh-in-situ: Movement and unselective binding. In *The Representation of (In)Definiteness*. Edited by Eric J. Reuland and Alice G. B. ter Meulen. Cambridge: MIT Press, pp. 98–129.
Poletto, Cecilia. 2000. *The Higher Functional Field. Evidence from Northern Italian Dialects*. Oxford and New York: Oxford University Press.
Poletto, Cecilia. 2002. The left-periphery of V2-Rhaetoromance dialects: A new view on V2 and V3. In *Syntactic microvariation*. Edited by Sjef Barbiers, Leonie Cornips and Susanne van der Kleij. Amsterdam: Meertens Instituut, pp. 214–42.
Poletto, Cecilia. 2005. Sì and e as CP Expletives in Old Italian. In *Grammaticalization and Parametric Variation*. Edited by Montserrat Batllori, Maria-Lluïsa Hernanz, Carme Picallo and Francesc Roca. Oxford: Oxford University Press, pp. 206–235.
Reinhart, Tanya. 1981. Pragmatics and linguistics: An analysis of sentence topics. *Philosophica* 27: 53–94. [CrossRef]
Renzi, Lorenzo, and Alvise Andreose. 2015. *Manuale di Linguistica e Filologia Romanza. Nuova Edizione*. Bologna: Il Mulino.
Renzi, Lorenzo, and Laura Vanelli. 1983. I pronomi soggetto in alcune varietà romanze. In *Scritti linguistici in onore di G.B. Pellegrini*. Edited by Paola Benincà and Giovan Battista Pellegrini. Pisa: Pacini, pp. 121–145.
Rizzi, Luigi. 1982. *Issues in Italian Syntax*. Berlin/Boston: De Gruyter Mouton.
Rizzi, Luigi. 1986. On the Status of Subject Clitics in Romance. In *Studies in Romance Linguistics*. Edited by Osvaldo Jaeggli and Carmén Silva-Corvalan. Dordrecht: Floris, pp. 137–52.
Rizzi, Luigi. 1990. *Relativized Minimality*. Cambridge: MIT Press.
Rizzi, Luigi. 1997. The fine structure of the left periphery. In *Elements of Grammar*. Edited by Liliane Haegeman. Dordrecht: Kluwer, pp. 281–337.
Rizzi, Luigi. 2013. Notes on cartography and further explanation. *Probus* 25: 97–226. [CrossRef]
Rizzi, Luigi. 2018. Uniqueness of left peripheral focus, "further explanation", and Int. In *Order and Structure in Syntax I: Word Order and Syntactic Structure*. Edited by Laura R. Bailey and Michelle Sheehan. Berlin: Language Science Press, pp. 333–43.
Rizzi, Luigi, and Ur Shlonsky. 2006. Satisfying the subject criterion by a non subject: English locative inversion and heavy NP shift. In *Phases of interpretation*. Edited by Mara Frascarelli. Berlin: Mouton de Gruyter, pp. 341–61.
Rizzi, Luigi, and Ur Shlonsky. 2007. Strategies of subject extraction. In *Interfaces + Recursion = Language? Chomsky's Minimalism and the View from Syntax-Semantics*. Edited by Hans-Martin Gärtner and Uli Sauerland. Berlin: Mouton de Gruyter, pp. 115–60.
Roberts, Ian. 2010. *Agreement and Head Movement: Clitics, Incorporation, and Defective Goals*. Cambridge, MA: MIT Press.
Rögnvaldsson, Eirikur, and Hoskuldur Thráinsson. 1990. On Icelandic word order once more. In *Syntax and Semantics: Modern Icelandic Syntax*. Edited by Joan Maling and Annie Zaenen. San Diego: Academic Press, pp. 3–40.
Rohlfs, Gerhard. 2021. *Grammatica storica della lingua italiana e dei suoi dialetti*. Bologna: Il Mulino. First published in 1968.
Sasse, Hans-Jürgen. 1987. The thetic/categorical distinction revisited. *Linguistics* 25: 511–80. [CrossRef]
Schaefer, Silvia. 2020. The morpho-syntactic encoding of discourse-linked topics: An agreement alternation in inversion in North-Eastern Italian varieties. *Glossa: A Journal of General Linguistics* 5: 96. [CrossRef]
Sorace, Antonella. 2000. Gradients in auxiliary selection with intransitive verbs. *Language* 76: 859–90. [CrossRef]
Sornicola, Rosanna. 1979. Egocentric reference as a problem for the theory of communication. *Journal of Italian Linguistics* 2: 7–63.
Sornicola, Rosanna. 1996. Alcune strutture con pronome espletivo nei dialetti italiani meridionali. In *Italiano e dialetti nel tempo. Saggi di grammatica per Giulio C. Lepschy*. Edited by Paola Benincà, Guglielmo Cinque, Tullio De Mauro and Nigel Vincent. Roma: Bulzoni, pp. 323–40.
Sornicola, Rosanna. 2010. La rappresentazione delle strutture locativo-esistenziali in un corpus di italiano. Uno studio sull'analizzabilità strutturale del discorso parlato. *Vox Romanica* 69: 111–40.
Stalnaker, Robert. 1974. Pragmatic Presuppositions. In *Semantics and Philosophy*. Edited by Milton K. Munitz and Peter K. Unger. New York: New York University Press, pp. 701–21.
Vikner, Sten. 1995. *Verb Movement and Expletive Subjects in the Germanic Languages*. Oxford and New York: Oxford University Press.
Vitolo, Giuseppe. 2006. L'uso del dimostrativo espletivo in contesto di frasi principale e dipendenti nella parlata salernitana. *Quaderni del Dipartimento di Linguistica, Università di Firenze* 16: 61–67.
Williams, Alexander. 2001. Null subjects in Middle English existential. In *Diachronic Syntax: Models and Mechanisms*. Edited by Susan Pintzuk, George Tsoulas and Anthony Warner. Oxford: OUP, pp. 164–187.
Zamboni, Alberto. 1998. Dal latino tardo al romanzo arcaio: Aspetti diacronico-tipologici della flessione nominale. In *Sintassi storica. Atti del XXX congresso internazionale della società di linguistica italiana*. Edited by Paolo Ramat and Elisa Roma. Rome: Bulzoni, pp. 127–46.

languages

MDPI

Article

Null Subjects in Non-Pro-Drop Languages: The Lens on French

Mara Frascarelli * and Giorgio Carella *

Department of Foreign Languages, Literatures and Cultures, University of Roma Tre, 00146 Roma, Italy
* Correspondence: mara.frascarelli@uniroma3.it (M.F.); giorgio.carella@uniroma3.it (G.C.)

Abstract: The contrast between languages such as Italian that allow subjects of tensed sentences to be null (i.e., pro-drop languages) and those like French that do not (i.e., non-pro-drop languages) is a classic issue for comparative syntactic research. Nevertheless, while several studies have been dedicated to pro-drop languages, distinguishing across different types, subject omission in non-pro-drop languages is generally misjudged as a marginal or substandard phenomenon. However, a more careful examination reveals that the occurrence of Null Subjects (NSs) in non-pro-drop languages is associated with distinct semantic and discourse imports. Based on a systematic corpus analysis, this work will confirm that NSs do occur in Colloquial French, especially in the case of expletive subjects. Furthermore, evidence will be provided for a crucial connection between subject omission, expletive types, and the morpho-syntactic categories of person/number for argument pronouns. This pilot work can thus open new perspectives for future research.

Keywords: null subject; pro-drop parameter; discourse categories; expletives; argument roles; phi-features

1. Introduction

The contrast between languages that allow subjects of tensed sentences to be null (like Italian, Greek, and Turkish) and those that do not (like English and French) is a classic issue for comparative syntactic research. Within the formal framework of generative grammar, the licensing of Null Subjects (henceforth, NSs) has been a major topic since the 1980s, both for its theoretical import and its connection with a parameter (cf. Jaeggli and Safir 1989 and Rizzi 1982). Indeed, several works have been dedicated to the licensing, acquisition, and interpretation of NSs (cf. Tomioka 2003; Holmberg 2005; Holmberg et al. 2009; Neeleman and Szendröi 2007; Frascarelli 2007; and Biberauer et al. 2010, among others), and different sub-types have been identified and approached.

On the other hand, few works have been dedicated to the occurrence and formal/discourse properties of NSs in languages which do not allow for subject omission in tensed sentences, as this phenomenon is generally considered marginal, substandard, or simply ungrammatical, and as such, somehow non-existent. In particular, up to the 1990s, two dominant hypotheses can be found: in one case the use of NSs is tied to (not clearly specified) 'social attitudes' (cf. Langacker 1985) while others believe that NSs occur in conversation for reasons of 'temporal efficiency' (Napoli 1982). However, though it is trivially true that deleting the subject shortens an utterance, no study could demonstrate that such efficiency is the point of NS realizations. On the contrary, later studies have shown that the shorter the conversational turn, the greater the possibility of the subject being null (cf. Section 6.3). Then, recent works (cf. Cote 1996; Haegeman 2000; Torres Cacoullos and Travis 2014; and Wagner 2012, among others) have shown that though restricted, the occurrence of NSs in non-pro-drop English is not random or the result of careless wording, but tightly connected with specific formal and discourse properties. As for French, several studies have shown that it allows the realization of NSs in some specific structural contexts and with a limited number of verbs, especially in impersonal constructions (cf. Zimmermann and Kaiser 2014 and Zimmermann 2018, among others).

Citation: Frascarelli, Mara, and Giorgio Carella. 2024. Null Subjects in Non-Pro-Drop Languages: The Lens on French. *Languages* 9: 363. https://doi.org/10.3390/languages9120363

Academic Editors: Ana Ojea and Jason Kandybowicz

Received: 7 May 2024
Revised: 28 October 2024
Accepted: 11 November 2024
Published: 27 November 2024

The present work aims at verifying the previous literature on the use of NSs in Modern Colloquial French, also offering a useful cross-linguistic comparison for future research in this area.

In particular, the present investigation is based on the analysis of three conversations extracted from two online corpora of Colloquial French, namely the CLAPI (http://clapi.icar.cnrs.fr/ accessed on 1 December 2023) and the CFPP2000 (http://cfpp2000.univ-paris3.fr/ accessed on 1 December 2023), for a total of two and a half hours of conversations between friends on different subjects (nine speakers and approximately 56,000 words).

In detail, the present paper is organized as follows: Section 2 is dedicated to the notion of pro-drop, a brief description of the NS [parameter], and a background on the studies dedicated to the interpretation of pro, whereas Section 3 presents a typology of pro-drop languages, distinguishing between consistent, partial, radical, and expletive pro-drop languages, with relevant examples from different languages. In Section 4 the lens of analysis concentrates on French, in order to present the results of previous studies, as well as providing elements of diachronic analysis that may be useful for the interpretation of phenomena found in Modern French. Section 5 introduces the research questions and illustrates materials and methodology. Then, Section 6 is dedicated to the quantitative and qualitative analysis of data and to a discussion on the most relevant trends. Finally, Section 7 is dedicated to the final conclusions and considerations for future research.

2. Null Subjects and the Null Subject Parameter

An NS is a subject that is not overtly expressed in a sentence and thus NS languages (NSLs) are those languages in which sentences can still be grammatical even if the subject is not explicitly pronounced. Conversely, in non-NSLs subjects must always be overtly realized (except in the case of imperatives, subject relatives, and some fixed expressions). For instance, in Italian, which is a quite consistent NS language, the subject can be silent in virtually any clause type context (as is shown in Frascarelli 2007 and related works), which is different from a non-NS language like French, as is shown in (1) and (2) below (cf. English as well in the translations):

(1)	*Leo*$_k$	*ha*	*detto*	*che*	*pro*$_k$	*uscirà*	*quando*
	Leo	have.3SG	say.PRT	that	NS	go out.FUT.3SG	when
	pro$_k$	*avrà*	*finito*	*il*	*suo*	*lavoro*	
	NS	have.FUT.3SG	finish.PRT	the	POSS.3SG	work	
	'Leo said that he will go out when he would have finished his work.'						
(2)	*Leo*$_k$	*a*	*dit*	*qu'*	***il***$_k$	*viendrait*	*lorsqu'*
	Leo	have.3SG	say.PRT	that	pron.3SGM	go out.FUT.3SG	when
	il$_k$	*aura*	*terminé*	*son*	*travail*		
	pron.3SGM	have.FUT.3SG	finish.PRT	POSS.3SG	POSS.3SG		
	'Leo said that he will go out when he would have finished his work.'						

Example (3) below also shows that in Italian (and in consistent NS languages in general) the person/number specification of the subject is irrelevant, as well as the tense/mood characterization of the verb:

(3)	*pro*$_z$	*penso*	*che*	*pro*$_k$	*andrà*	via	*sebbene*
	NS	think.1SG	that	NS	go.FUT.3SG	away	though
	pro$_w$	*avrebbero*	*preferito*	*che*	*rimanesse*	*qui*	
	NS	have.COND.3PL	prerfer.PRT	that	stay SUBJ.3SG	here	
	'I think he will leave although they would have preferred him to stay here.'						

Of course, the subject of a clause in an NS language can also be overt. With an overt subject, the Italian sentences above are indeed perfectly grammatical. However, it has been argued that the overt realization of subjects in pro-drop languages is associated to discourse-related requirements; in other words, this means that an overt subject in a language like Italian is either a Focus or (some type of) Topic (as in, respectively, (4) and (5) below) (for discussion, cf. Grimshaw and Samek-Lodovici 1998; Frascarelli 2007, 2018; and Mayol 2010).

(4)	*Io*	*vado*	via	*(non*	*Sara)*
	pron.1SG	go.1SG	away	not	Sara
	'It's me who is going (not Sara).'				

(5)	*Io$_k$,*	*vado*	via	*(…gli*	*altri*	*pro$_k$*	*non*	*lo*	*so)*
	pron.1SG	go.1SG	away	the	others	NS	not	OBJ.CL.3SG	know.1SG
	'(as for me) I'm going (the others I don't know).'								

Rizzi's Null Subject Parameter (NSP)

Building on previous studies investigating the properties of NSLs and the licensing of NSs (cf. Perlmutter 1971; Chomsky and Lasnik 1977; Taraldsen 1980; Kayne 1980; and Chomsky 1981, among others), in his 1982 seminal work, Rizzi formulated the Null Subject Parameter (NSP). This paved the way for a vast amount of cross-linguistic research and analyses, leading to an understanding of various phenomena, but also highlighting the limits deriving by positing a strict connection between NS licensing and inflection.

Leaving aside the exposition of data concerning free subject–verb inversion, rich agreement, and the absence of "that-t" effects (the interested reader can refer to Perlmutter 1971; Kayne 1980; Holmberg 2005; and D'Alessandro 2015, among many others), non-referential (expletive), NSs will be briefly treated below, since they will play an important role in the corpus analysis conducted in this paper.

As is known, expletive subjects occur in all the contexts in which the subject has no argument role, and according to Rizzi's (1982) NSP, if a language has thematic NSs it must also have null expletives. This means that in pro-drop languages like Italian, expletive subjects are covert, whereas in non-pro-drop languages non-referential subjects must be overt, like referential ones.[1] Consider the following contrasts between Italian and French (and English, in the translations):[2]

(6)	pro$_{EXPL}$	*sembra*	*(che)*	pro$_{EXPL}$	*sia*	*tardi*
	EXPL	seem.3SG	that	EXPL	be.SUBJ.3SG	late
(7)	*il$_{EXPL}$*	*semble*	*qu'*	*il$_{EXPL}$*	*soit*	*tard*
	EXPL	seem.3SG		EXPL	be.SUBJ.3SG	late
	'It seems (that) it is late.'					

(8)	pro$_{EXPL}$	*sta*	*piovendo*
	EXPL	be.3SG	raining
(9)	*il$_{EXPL}$*	*pleut*	
	EXPL	rain.3SG	
	'It is raining.'		

(10)	pro$_{EXPL}$	*è*	*probabile*	*che*	*ci*	*sia*	*sciopero*	*domani*	
	EXPL	be.3SG	probable	that	there	be.SUBJ.3SG	strike	tomorrow	
(11)	*il$_{EXPL}$*	*est*	*probable*	*qu'*	*une*	*grève*	*ait*	*lieu*	*demain*
	EXPL	be.3SG	probable	that	a	strike	have.SUBJ.3SG	place	tomorrow
	'There's likely to be a strike tomorrow.'								

(12)	pro$_{EXPL}$	*bisogna*	*sbrigarsi*
	EXPL	must.3SG	hurry
(13)	*il$_{EXPL}$*	*faut*	*se dépêcher*

EXPL must.3SG hurry
'One/you must hurry.'

(14) pro$_{EXPL}$ *c'* *è* *un* *uomo* *in* *giardino*
EXPL there be.3SG a man. in garden

(15) *il*$_{EXPL}$ *y* *a* *un* *homme* *dans* *le* *jardin*
EXPL there have.3SG a man iiin the garden
'There is a man in the garden.'

(16) pro$_{EXPL}$ *è* *incredibile* *ma* *vero!*
EXPL be.3SG incredible but true

(17) *c'*$_{EXPL}$ *est* *incroyable* *mais* *vrai*
EXPL be.3SG incredible but true
'It's incredible but true.'

Comparing Italian with non-pro-drop languages, such as English and French, Kayne (1980) also observes that the latter must have an expletive in the canonical subject position in the case of so-called subject inversion, as is shown in (18) and (19) and the corresponding English translation.

(18) è Leo che lo sa
be.3SG Leo that it.OD know.3SG

(19) *(c') *est Leo qui (le) sait
it be.3SG Leo who it.OD know.3SG
'*(it) is Leo who knows.'

By looking at the data presented so far, we might be tempted to conclude that the NS status of a language is a pretty straightforward matter. You must express the subject, like in English/French, or you do not have to, like in Italian/Spanish. However, empirical evidence shows that both pure pro-drop and non-pro-drop languages are rare. In fact, most languages only feature NSs of a specific type and/or in specific structural contexts, as will be shown in the following section.

3. A Typology of Pro-Drop Languages

As we have seen so far, pro-drop languages are those languages that allow for a covert realization of the subject of a sentence. However, as mentioned at the end of Section 2, pro-drop languages do not behave consistently, and thus, they have been classified in different groups based on the nature of the NSs allowed and/or the specific contexts in which NSs can occur.

3.1. Consistent Pro-Drop Languages

As the name suggests, consistent (or 'full') pro-drop languages are those languages which present the full range of characteristics included in Rizzi's NSP, namely free subject-verb inversion, rich agreement, and the absence of "that-t" effects. For this reason, they are also referred to as 'canonical' pro-drop languages. Consistent pro-drop languages include Arabic, Basque, Berber, Greek, Hausa, Turkish, and all romance languages except French.

3.2. Radical Pro-Drop Languages

Radical pro-drop languages, also referred to as 'discourse' pro-drop languages, are those which present the same characteristics as full pro-drop languages, but they lack any verbal inflection.[3] This type includes many Asian languages, such as Chinese, Japanese, Korean, Thai, and Vietnamese. The specificity of these languages is that grammar does not offer any help in identifying the covert subject, due to the lack of verb–subject agreement. Thus, it has to be retrieved from the discourse, as can be seen in the following examples (Adapted from Huang 1984, p. 533):

(20)	Speaker A:	*Zhangsan*	*kanjian*	*Lisi*	*le*	*ma?*	[Chinese]
		Zhangsan	see	Lisi	LE	Q	
		'Did Zhangsan see Lisi?'					
	Speaker B:	pro	*kanjian*	*ta*	*le*		
		NS	see	he	LE		
		'['[He] saw him'					

The licensing of *pro* in radical pro-drop languages has been the object of many studies in the last decades and many proposals have been put forth. For instance, Huang (1984) argues that in discourse-oriented languages such as Chinese (cf. Tsao 1977), empty variables like NSs are bound to a 'zero-topic', which licenses the *pro*. Saito (2007) argues for "a mechanism of PF merging of arguments that are copied directly from discourse elements" (D'Alessandro 2015, p. 221). Then, Barbosa (2011a, 2011b) proposes that NSs are in fact a case of (null) NP anaphora, while for Duguine (2013) they are instead NP/DP ellipsis. Finally, in Frascarelli and Casentini (2019), a discourse-related approach is assumed and an Agree relation is proposed between NSs and a dedicated type of Topic, namely the *Aboutness-shift* (A)-Topic; radical NS languages are thus treated in line with Frascarelli's (2007) analysis of consistent pro-drop languages, although with specific restrictions.[4]

Though interesting, a more in-depth analysis of the implications triggered by different approaches goes far beyond the scope of the present paper; the interested reader is thus referred to the studies mentioned above.

3.3. Partial Pro-Drop Languages

Partial pro-drop languages are those languages in which referential subjects can be covertly realized only in some structural contexts or based on their specific features. These languages include Finnish, Marathi, Russian, Icelandic, Assamese, Hebrew, and Brazilian Portuguese (cf. Biberauer et al. 2010). For instance, Finnish only allows for first and second person subjects to be null, while omitting a third person subject yields ungrammatical sentences, as shown in (21).

(21)					
a.	*(Minä)	*puhu-n*	*englantia*		[Finnish]
	I	speak-1SG	English		
b.	*(Sinä)	*puhu-t*	*englantia*		
	You	speak-2SG	English		
c.	*(Hän)	*puhu-u*	*englantia*		
	He/She	speak-3SG	English		
d.	*(Me)	*puhu-t*	*englantia*		
	We	speak-1PL	English		
e.	*(Te)	*puhu-tte*	*englantia*		
	You	speak-2PL	English		
f.	*(He)	*puhu-vat*	*englantia*		
	They	speak-3PL	English		

(Adapted from Holmberg 2005, p. 539)

On the other hand, Holmberg (2005) shows that Finnish third person pronouns can be covert if they are bound by an overt DP, as shown in (22). In this respect, Ylinärä and Frascarelli (2021) show, in turn, that Holmberg's 'binding DP' is in fact an A-Topic, which is in line with Frascarelli's (2018) proposal for partial pro-drop languages.

(22)	*Pekka*$_i$	*väittää*	*[että*	*hän*$_{i/j}$ */*$\varnothing_{i/*j}$	*puhuu*	*englantia*	*hyvin]*	[Finnish]
	Pekka	claims	that	he	speaks	English	well	

(Holmberg 2005, p. 539)

It should be finally noticed that the possibility to covertly realize a subject in partial pro-drop languages can be restricted in the presence of a generic/indefinite subject, as is the case of Marathi.

(23) *Unahlyat* *lavkar* *utthavla* *jato* [Marathi]
summer-in early wake go-PRES.3SG.M
'In summer one wakes up early'
(Holmberg et al. 2009, p. 125)

3.4. Expletive Pro-Drop Languages

Finally, there are the so-called "Expletive NS languages", which only allow for expletive subjects to be null. This is the case for Dutch (Gilligan 1987):

(24) *Gisteren* *werd* *(er)* *door* *het* *hele* *dorp* *gedanst* [Dutch]
yesterday be.PST.3SG there by the whole village dance.PST.3SG
'Yesterday, there was dancing by the whole village'

As can be seen in (24), the expletive pronoun *er* can be omitted in the embedded sentence. Finnish, a partial NS language, also features some null expletives as shown in (25) (Holmberg and Nikanne 2002). (For a discussion on different expletive types and their properties, also cf. Ylinärä and Frascarelli 2021).

(25) *Nyt* *(se)* *taas* *sataa* [Finnish]
now EXPL again rain.PST.3SG
'Now it's raining again

4. Subject Omission in French: A Short Historical Overview

As said in the introduction, the present paper is primarily intended to verify the results of previous studies on the use of NSs in Modern Colloquial French. Nevertheless, since the pro-drop quality has different implementations (as shown in Section 3) and can vary over time (giving rise to 'cyclic variations', cf. Givón 1976), we consider it useful to check whether French has ever admitted subject omission of some kind in the past. Before turning to a corpus analysis of contemporary spoken data, this section is therefore dedicated to a brief preliminary excursus of the relevant literature concerning Old French, as well as to a brief report on the results obtained by previous research on the use of NSs in Modern French.

4.1. NSs in Old French

A long-standing problem in Old French syntax is the ability to account for the variability in subject–pronoun realization. As can be seen from the examples below, taken from early 13th century narrative prose, an overt subject pronoun could appear either before the tensed verb (26), or after it (27), if a non-subject constituent stood in first position, whereas an NS occurred only in the latter position.

(26) *Ele* *vint* *as* *deus* *rois*
she come.PST to.DEF.PL two king.PL
'She came to the two kings.' (*La mort le roi Artu*)

(27) *Et* *por* *ce* *vos* *vint* *ele* *veoir*
and for this you.OBJ come.PST she see.INF
'And therefore she came to see you.' (*La queste del saint Graal*)

In this line of research, Vance (1997) noticed that in prose romances first and second person pronouns, which are generally used in dialogues, were more often overt, whereas third person subjects, most commonly used in narrative parts, were more often null.

Within contemporary syntactic theory the question tends to be whether Old French was a pro-drop language typologically. Roberts (1993) and Rinke (2003) considered that it was, whereas Zimmermann (2014) argued to the contrary. However, the question should not be seen as a binary choice, since as discussed in Section 3.3 above, research has demonstrated the existence of *partial* pro-drop languages such as Finnish and Brazilian Portuguese (Holmberg and Nikanne 2002; Modesto 2008; Ylinärä and Frascarelli 2021), in which subjects may be left null in restricted contexts.

In this respect, Ingham's (2018) corpus study provides a crucial contribution for the understanding of subject omission in Old French, since it shows that definite third person NSs are allowed only when they are coreferential with the subject–Topic of a higher clause and that NSs typically appear in chains. Ingham thus shows that these chains are headed by an element that qualifies as what Frascarelli (2007) defines the A-Topic (cf. note 4). The A-Topic was later refined by Holmberg (2010) in a Minimalist approach, by postulating that Frascarelli's Agree relation implies the evaluation of the D-feature of the NS (inherited from T), which is thus provided with a referential index.

Furthermore, Ingham (2018) notices that the realization of NSs in Old French seems to be limited to root clauses and is associated with the appearance of the particle *si*, which is used to introduce to the content of the preceding clause, similarly to its Latin etymon *sic* ('so'). Since similar constraints do not apply in consistent pro-drop languages, the author proposes a "partial" pro-drop characterization for Old French.

As a matter of fact, the status of NSs in Old French is generally linked to the validity of the V2 analysis of root clauses (Rinke 2007). However, while Zimmermann (2018) only focused on expressions involving V-movement to the left periphery, Ingham argued that both left-dislocated Topics and Focus elements produce V2 structures, with V moving to Fin^0 in both cases. In this line of analysis, it can be feasibly proposed that the preposing of the particle *si* allowed for subject omission.

(28)	*Et*	*quant*	*Phariens*	*voit*	*ce,*	***si***	*saisit*
	and	when	Phariens	see.PRES.3G	this	SI	seize.PRES.3G
	une	*hache*	*qu'*	*il*	*avoit*	*en*	*la*
	an	axe	that	he	have.IMPF.3SG	in	the
	tor	*maint*	*jor*	*gardee,*	***si***	*s'escorce*	
	tour	many	days	keep.PP.FEM	SI	REFL.rush.PRES.3SG	
	vers	*son*	*niveau*				
	towards	is	nephew				
	'And when P. sees this, he seizes an axe which he had kept in the tower for a long time, and rushes towards his nephew.'						(*Lancelot du Lac.* Tome I)
(29)	*Et*	*ele*	*le*	*regarde,*	***si***	*le*	*conoist*
	and	she	him	look-at.PRES.3SGSI		him	know.PRES.3SG
	'And she looks at him, recognises him.'						(*La queste del saint Graal*)

As for non-argument subjects, the analysis of medieval prose texts shows that Old French expletive pronouns occurred in all types of constructions in which a thematic/referential subject was not available, that is to say, in constructions with so-called quasi-arguments as subjects and in constructions with non-arguments in subject position. Expletives occurred predominantly (but not exclusively) in embedded clauses and occasionally post-verbally in matrix clauses. Consider examples (30) to (34) below:

(30) *…que* *il* *ne* *plúve* *pur* *lur* *pecchié…*
because it not rain.PRES.3SG for their sins
''… because it does not rain on account of their sin…' (*Livre Reis* 3, 8, 35)

(31) *…se* *il* *te* *semble,* *au* *tuen* *avis,* *que…*
if it to.you seem.PRES.3SG to.DET your opinion that
''… if it seems to you, in your opinion, that …' (*Saint Graal* 510, 37)

(32) *Quant* *ce* *vint* *le* *soir…*
when it come.PST.3SG the evening…
''When it was dark, ...' (*Saint Louis* 410)

(33) *Il* *nen* *I* *ad* *chevalier* *ne* *barun…*
it not there have.IMPF.3SG knight nor baron
''There was no knight nor baron...' (*Roland* 2418)

Through the analysis of Old French poems and prose texts from the IX to the XIV century, Zimmermann (2009) argues against an analysis of Old French as a V2 language, showing that the realization of subject pronouns was not contingent upon the V2 constraint. Importantly, Zimmermann shows that the realization of expletive pronouns in the prose texts selected increases, though not linearly, both in matrix and embedded clauses, and the same kind of evolution can be seen for thematic subjects and demonstratives. This increasing trend is at its maximum in the 14th century texts, leading the author to analyze Old French as a non-pro-drop language and NSs in Old French as relics of an earlier language stage, which could (still) be realized as long as certain specific structural conditions were met.

A full discussion of expletive subjects in Old French is far beyond the scope of this paper. What is important to underline is that null expletive subjects have been present in the French language since the Middle Ages, and as a consequence, the non-pro-drop quality of Modern French includes both argument and non-argument subjects.

4.2. NSs in Modern French

The use of NSs in Modern French has been examined in a range of studies, revealing that while French is typically categorized as a non-pro-drop language, NSs are still possible and at times even obligatory in specific contexts. Zimmermann (2018) provides a detailed overview of these contexts, indicating that prosodically weak (referential and expletive) subject pronouns can be omitted in specific environments. These include, among others, (i) comparative clauses, (ii) clauses with left-dislocated prosodically strong pronouns, (iii) imperatives and stylistic inversion in root and embedded interrogatives, and (iv) embedded subjunctives (for a detailed discussion, see Zimmermann 2018). Additionally, certain impersonal clauses allow NSs, particularly with verbs like *rester* 'to remain' and *suffire* 'to be sufficient', as well as with a limited class of passive verbs in specific syntactic configurations. The author further notes that NSs may appear more freely in particular written registers of Modern French that prioritize linguistic economy, such as diaries, text messages, and report cards.

Zimmermann and Kaiser (2014) provide instead a closer examination of the conditions under which the expletive *il* may be omitted in Colloquial French. Their analysis highlights that the omission of *il* is most frequent with a select set of impersonal verbs, such as *y avoir* 'to exist' (i.e., existential constructions) and *falloir* 'to have to', which are more likely to appear without the overt expression of *il*. Other verbs like *valoir mieux* 'to be better' and *faire* 'to make' + adjective (+infinitive) exhibit a lower frequency of null *il*, while verbs such as *paraître* 'to appear', *sembler* 'to seem', and *suffire* 'to be sufficient' show the least frequent occurrences of null *il*.

Together, these studies illustrate that while French generally adheres to a non-pro-drop pattern; the presence of NSs in certain syntactic and stylistic contexts reflects a more nuanced distribution. The present work aims at verifying these results, as we will describe in detail in the following section.

5. Materials, Methods, and Goals

Given the above background, the question is whether the results reported in Section 4.2, coming from different studies and based on different methodologies, can be confirmed by a systematic and comprehensive corpus analysis. In addition, we considered the significant differences between non-pro-drop languages attested in the literature, and a cross-linguistic comparison seems to be in order. Hence, the main goal of the present study is to provide an answer to the following research questions:

1. What types of subjects (and how often) are indeed allowed to be omitted in Modern Colloquial French?
2. Does the use of NSs in French have patterns similar to other non-pro-drop languages such as English?

In order to accomplish this objective, we will discuss the results of a corpus analysis of spoken data in face-to-face spontaneous conversations taken from two spoken French corpora:

a. The CFPP2000 corpus (http://cfpp2000.univ-paris3.fr/index.html accessed on 1 December 2023), containing 58 interviews collected after 2000 on the territory of Paris and the nearby suburbs. This is a very well-organized corpus, providing very detailed descriptions of speakers.
b. The CLAPI corpus, containing 194 transcripts documenting oral language in different types of linguistic situations (informal, professional, institutional, commercial, etc.; http://clapi.icar.cnrs.fr accessed on 1 December 2023).

These corpora have been examined with the collaboration of two senior students who devoted about 200 h to this work. The transcripts used for the analysis have been selected based on the following criteria: (a) informal communicative situations, so as to favor naturalness; (b) significant presence of long conversational turns, so as to avoid interruptions and have the possibility of examining the realization of subjects both in matrix and subordinate clauses; (c) medium–high education of the speakers, in an effort not to associate subject omission to diastratic factors; (d) speakers aged between 18 and 30–35, to focus on the age group most open to linguistic changes; and finally, (e) recordings occurred from 2000 onward, in order to examine the current situation.

Three transcripts were thus selected: (1) "*Auréane L'huisser et Pierre-Fabien Benoît* " from the CFPP2000 corpus (http://cfpp2000.univ-paris3.fr/Corpus.html accessed on 1 December 2023) and (2) "*Apéritif entre ami(e)s—chat*" (http://clapi.icar.cnrs.fr/V3_Feuilleter.php?num_corpus=107 accessed on 1 December 2023) and (3) "*Montage meuble*" (http://clapi.icar.cnrs.fr/V3_Feuilleter.php?num_corpus=102 accessed on 1 December 2023), both from the CLAPI corpus. For the sake of convenience, in the rest of the article we will refer to these three conversations as, respectively, "*Auréane*", "*Montage*" and "*Apéritif*".

Auréane is a one-and-a-half hour interview with three speakers: two young people in their 20s from the 18th arrondissement of Paris and the interviewer who is about 50 years old. *Apéritif* portrays a friendly conversation of about half an hour between four young men, also in their 20s. Finally, *Montage* records the conversation, also half an hour long, of two girls in their 20s struggling with the assembly of an Ikea piece of furniture. The sub-corpus thus created for this study consists of two and a half hours of conversation for a total of 56,272 words (specifically, 25,709 from *Auréane*, 15,440 from *Apéritif*, and 15,123 from *Montage*) and 4996 sentences (3179 from *Auréane*, 956 from *Apéritif*, and 861 from *Montage*).

Then, fearing shortcomings or "normalizations", we proceeded to listen carefully to the audios and check the fidelity and accuracy of the transcriptions against the speech. This fear proved to be well founded, as listening to these passages revealed several instances in which the transcribers had not remained entirely faithful in their recording of speech, making arbitrary corrections or omitting elements.

Thus, several other cases of NSs have been detected, both expletive and referential, although the latter only in restricted contexts (cf. Section 6.2). Significant phenomena of the "morphological collapse" of pronouns were also noticed, especially regarding the first singular pronoun *je* (cf. Section 6.4). The data found have been reported in Excel files and categorized in detail according to the different types, also reporting their contexts of occurrence.

The results have been statistically analyzed, primarily through the Chi-squared test, for a rigorous examination of the frequency distributions within the data, particularly when comparing the occurrence of null and overt subjects across different syntactic constructions. Additionally, due to the presence of small sample sizes in certain instances, the Fisher Exact test has been utilized to ensure robust statistical inference. Finally, when necessary, a z-test for one proportion has been employed to further investigate specific patterns or trends within the dataset.

6. Results and Discussion

As mentioned in Section 5, the analyzed sub-corpus contains a total of 4996 sentences. Of these, 4600 (92.07%) feature an overt subject, while only in 396 cases (7.93%) the subject has been left unexpressed. This result confirms what has been found in previous studies (cf. Section 4.2), namely, that although French is classified as a non-pro-drop language, there are certain contexts in which NSs can occur. More specifically, results vary significantly among the three transcripts, as shown in Table 1.

Table 1. Null vs. overt Subjects.

Transcript	NULL		OVERT		Tot
	#	%	#	%	
Auréane	294	9.25	2885	90.75	**3179**
Apéritif	49	5.13	907	94.87	**956**
Montage	53	6.16	808	93.84	**861**
TOT	396	7.93	4600	92.07	**4996**

As can be seen in Table 1, the first transcript presents a significantly higher proportion of NSs compared to both the second ($X^2 = 16.4$, $p = < 0.001$) and the third ($X^2 = 8.2$, $p = 0.004$). This (relatively small) difference between the occurrence of NSs in the three transcripts could suggest a potential inclination towards a more elliptical or context-dependent speech by the speakers of the first conversation compared to those of the other two. Conversely the rate of NSs in the other two transcripts is not significantly different ($X^2 = 0.9$, $p = 0.341$).

As it will be shown in the next subsections, the semantic nature of the subjects seems to be a key factor in determining the likelihood of their covert realization. Indeed, 367 (92.68%) of the 399 identified NSs are expletives, while only 29 (7.32%) are referential.

6.1. Expletive NSs

The analysis of the sub-corpus showed that 1789 sentences out of 4996 (35.81%) feature an expletive subject, and as mentioned above, the absolute majority of the NSs are expletives. Indeed, null expletives constitute 20.51% of the total expletive subjects, while only 0.9% of the referential subjects are left unexpressed.

Specifically, null expletives have been identified in different syntactic contexts, namely existential (*il*) *y a* constructions (34a), impersonal (*il*) *faut* and (*il*) *arrive* sentences (34b–c), light verb structures featuring (*il*) *fait* (34d), and fixed expressions such as *s'*(*il*) *vous plaît* (34e)[5] and (*il*) *vaut mieux* (34f):

(34) a. *(Il)* *y* *a* *plein* *de* *gens*
EXPL there have.3SG plenty of people
'There are plenty of people' (*Auréane*)

b. *(Il)* *faut* *qu'* *on* *y* *aille*
EXPL need. 3SG that one there go.SUBJ.3SG
'We need to go there' (*Apéritif*)

c. *(Il)* *peut* *m'* *arriver* *de* *m'* *adonner*
EXPL may.3SG to.me happen.INF of my.self devote.INF
à *d'* *autres* *produits*
to of other products
'I may occasionally indulge in other products' (*Auréane*)

d. *(Il)* *faisait* *vraiment* *chaud*
EXPL make.PST.3SG really hot
'It was really hot' (*Apéritif*)

e. *T'* *enlève* *ça* *s'(il) te plaît*
PRON.DAT.2SG remove.IMP.2SG that please
'Please take it off' (*Montage*)

f. *(Il)* *vaut* *mieux* *avoir* *l'* *air* *de*
EXPL be.worth.3SG better have. INF DET air of
bien *vouloir* *engager* *la* *conversation*
well want. INF engage. INF det conversation
'It's better to look like you want to start the conversation' (*Apéritif*)

Furthermore, it is important to note that these structures have varying rates of occurrence within the sub-corpus examined, as reported in Table 2.

Table 2. Occurrence rate of expletive structures.

Structure	#	%
(il) y a	380	79
(il) faut	78	16.22
(il) arrive	3	0.62
(il) fait	13	2.7
s'(il) vous plaît	4	0.84
(il) vaut mieux	3	0.62
TOT	**481**	**100**

Data show that existential (*il*) *y a* constructions constitute the great majority, accounting for 79% of the occurrences. Impersonal expressions featuring (*il*) *faut* follow with 16.22%, while sentences with (*il*) *fait* represent a smaller proportion at 2.7%. Light verb constructions with (*il*) *arrive* and fixed expressions such as *s'*(*il*) *vous plaît* and (*il*) *vaut mieux* exhibit minimal occurrences in the dataset. These varying rates of occurrence underscore the differential distribution and usage patterns of expletive structures in French colloquial speech, reflecting the diverse linguistic contexts in which they appear.

Turning to the analysis of NSs across these syntactic constructions, results show significant variations in their frequency and distribution, as shown in Table 3.

Table 3. Null vs. overt expletives.

	NULL		OVERT	
	#	%	#	%
(il) y a	315	82.89	65	17.11
(il) faut	42	53.85	36	46.15
(il) arrive	2	66.67	1	33.33
(il) fait	2	15.38	11	84.62
s'(il) vous plaît	3	75.00	1	25.00
(il) vaut mieux	3	100.00	0	0.00

As can be seen in Table 3, speakers tend to realize most Null Subjects in existential (*il*) *y a* constructions (82.89%). Conversely, impersonal (*il*) *faut* sentences show a more balanced distribution, with no significant difference between null and overt realizations ($z = -0.68$, $p = 0.496$). Interestingly, light verb structures featuring (*il*) *fait* display a stark contrast, with NSs representing only 15.38% of instances. Finally, impersonal (*il*) *arrive* sentences and fixed expressions such as *s'*(*il*) *vous plaît* and (*il*) *vaut mieux* present no significant differences with respect to (*il*) *y a* constructions (a Fisher's Exact test yielded, respectively, $p = 0.433$, $p = 0.531$, and $p = 1$). However, it is important to note that the results concerning these three structures should be interpreted cautiously due the already mentioned small sample size, which may impact the reliability and generalizability of the findings and warrant further investigation to validate these trends.

On the one hand, the results presented so far are in line with what has been found in previous studies (cf. Section 4.1), in which *(il) y a* and (*il*) *faut* constructions appear to be the most likely to occur with an NS, while other verbs present a lower frequency of NSs. On the other hand, our analysis identified a few instances of NSs occurring with verbs which, as far as we know, have not been mentioned in the literature. This seems to suggest that a definitive list of specific verbs allowing an NS in Colloquial French is far from being completed (if it ever will be) and that it might be more fruitful to look at the specific linguistic context which triggers the realization of an NS in a non-pro-drop language such as French.

In this respect, it is interesting to notice that the absolute majority of the expletive subjects identified in the corpus (1301, that is 72.72%) are those occurring in non-predicational copular sentences (Den Dikken 2006a), namely the *ce* in *c'est* constructions. Please consider the following examples:

(35) a. *C'* *est* *une* *amie* *à* *moi*
EXPL be.3SG DET.INDEF.F friend F to pron.ACC.1SG
qui *m'* *avait* *dit*
pron.REL.NOM.1SG CL.ACC.1SG have. PST.3SG tell. PP
'It's a friend of mine who told me….' (*Apéritif*)

b. *Mais* *c'* *est* *vraiment* *genial*
but EXPL be.3SG really brilliant
'But it's really brilliant!' (*Montage*)

c. *C'* *est* *une* *soupe* *populaire* *musulmane*
EXPL be.3SG DET.INDEF.F soup common Muslim
'It's a Muslim kitchen-soup.' (*Auréane*)

d. *C'* *est* *pour* *ça* *qu'* *J'*
EXPL be.3SG for that that pron.NOM.1SG
disais *le* *carton*
say.PST.IMPF.1SG DET.F box
'That's why I said the box.' (*Auréane*)

Interestingly, not a single one of this type of expletive has been omitted by speakers. This result may be explained by the fact that, while existential (*il*) *y a* constructions predicate the existence of the following entity, impersonal (*il*) *faut* sentences predicate a need, and (*il*) *arrive* sentences predicate possibility,[6] "*c'est*" constructions are non-predicational in nature. In short, the different behavior shown by the two French expletives, *il* and *ce*, seems to show that the crucial factor for the licensing of NSs in Colloquial French is whether an expression is predicative in nature or not.

This proposal is supported by recent analyses in which the different (information) structural functions of the relevant expletives have been highlighted. In particular, we refer to Frascarelli (2010a, 2010b) and Frascarelli and Ramaglia's (2013) works on (pseudo-) clefts and existential 'there' sentences.

Trying to briefly set out a long and complex argument, in the above-mentioned works Frascarelli and Ramaglia consider specificational sentences as copular sentences, based on Den Dikken's (2006a) influential proposal. As such, the structure of copular sentences implies a Small-Clause (SC) construction in which one of the two constituents specifies the value of the variable represented by the other. In this line of analysis, Belletti (2005) and Frascarelli (2010b) propose a monoclausal specificational study of cleft sentences in which the clefted phrase and the relative clause are merged as independent constituents within a SC. In particular, the clefted phrase is merged as the predicate, while the relative clause (i.e., presupposed information) can be assumed to play the subject role.

Hence, a cleft sentence can be described as a Focus-Presupposition structure (Krifka 2006), including a copular element, a focused constituent (i.e., the clefted phrase), and a subordinate clause.

(36) It is a book that I gave John
COP [$_{SC}$ [$_{DP}$ *that I gave John*] [$_{DP}$ *a book*]]

Frascarelli and Ramaglia (2013, 2014) took this proposal, and based on syntax–prosody interface evidence, proceeded a step further in the analysis of specificational sentences, showing that the relative clause should be analyzed *as a Topic*. Specifically, in line with Frascarelli and Hinterhölzl (2007), the authors assume that Topics must be distinguished according to their formal and discourse properties and that different types of Topics are located in dedicated functional projections in the (split) C-domain. As the relative clause in a cleft sentence is associated with a [+given] semantic property, it qualifies as a so-called 'Familiar Topic'. As such, it is subject to Merge in the lowest left-peripheral Topic position (FamP) and is realized with a deaccented prosodic contour. Its final right-peripheral position is derived through IP-inversion to the Spec position of the Ground Phrase (GP; cf. Poletto and Pollock 2004; for details, cf. Frascarelli 2007).

The derivation of a sentence like (38) is thus the following (please, note the position and the co-indexing of the expletive *it*, highlighted in bold):[7]

(37) a. [$_{GP}$ [$_{FocP}$ [$_{TopP}$ [$_{DP}$ OP$_k$ that I gave John e$_k$] $_z$ [$_{IP}$ is [$_{SC}$ **it$_z$** [$_{DP}$ a book]]]]]] →
b. [$_{GP}$ [$_{FocP}$ [$_{DP}$ a book]$_k$ [$_{TopP}$ [$_{DP}$ OP$_k$ that I gave John e$_k$]$_z$ [$_{IP}$ **it$_z$** is [$_{SC}$ t$_z$ t$_k$]]]]] →
c. [$_{GP}$ [$_{IP}$ **it$_z$** is [$_{SC}$ t$_z$ t$_k$]] [$_{FocP}$ [$_{DP}$ a book]$_k$ [$_{TopP}$ [$_{DP}$ OP$_k$ that I gave John e$_k$]$_z$ t$_{IP}$]]]

As can be seen, in this line of analysis the subject pronoun *it* in the SC *is not an expletive* but *a resumptive pronoun* of the right-hand topicalized relative DP, which is merged as the subject of the SC (39a) and moved to Spec, IP (39c). A final note concerns the copula, which, according to this approach, has no semantic content (also cf. Stowell 1981). It is just a functional element (i.e., a "linker"; cf. Den Dikken 2006b) of the two major constituents of the sentence, triggering the movement of its complement to the Spec position.

To conclude, according to this semantic and IS-approach that we assume, specificational constructions are realized as copular sentences in which new information serves as the predicate of a SC, and the presupposed part of the sentence is topicalized and resumed by a subject pseudo-expletive pronoun in the subject position.

Given this resumptive function, the subject of a specificational sentence (i.e., *ce* in French) *is not an expletive* but a referential pronoun, and as such, it can be hardly silent in a pro-drop language like French. This proposal provides a feasible explanation for the fact that the pronoun *ce* is always present in the corpora examined. As a matter of fact, even when a co-indexed Topic is apparently not realized, it is in fact present, albeit silently, as it can be deduced from the context (it can thus refer to something being talked about, as in the case of sentence (37b) above).

On the contrary, the expletive *il* in predicational sentence is indeed an expletive pronoun, which is merged in the subject position because no argument can move there to meet EPP requirements. In particular, this happens when the theme argument is propositional (hence, too 'heavy' to move in subject position, as with a raising verb like *seem* in English), or in presentational sentences, in which new information is a theme selected by the verb. This is exactly the case of the expletive *il* in *il y a* constructions in French.

Indeed, the French language has maintained the proto-Indo-European form of existential sentences, so that what is now generally realized through the auxiliary *be*, it is still realized in the original 'have+LOC' form (cf. Freeze 1992).

Following Den Dikken (2006a), the basic difference with respect to specificational sentences is that existential constructions are analysed as copular structures of the predicative type and, as such, characterized by the non-referentiality of the second nominal. The locative argument is realized in the VP through a cltitic pronoun (ci in Italian and y in French; also cf. La Fauci and Loporcaro 1997), which is co-indexed with the topicalized locative constituent. The expletive pronoun (*there* in English and *il* in French), does not consequently have an anaphorical function and is only inserted to meet the EPP requirement and predicate the *property* of the first nominal (its subject; cf. Ramaglia and Frascarelli 2019 for details).

The basic semantic and informational distinction characterizing expletives in specificational and predicational sentences can be thus feasibly assumed to be the explanation for their different behaviour and allow for a principled distinction of expletive drop in a non-pro-drop language like French.

As a final support to this proposal, it can be noticed that expressions such as *s'(il) vous plaît* and *(il) vaut mieux* can be considered examples of phrasemes, namely fixed expressions whose meaning is not directly derived from the meanings of their individual components. In these expressions, the expletive *il* does not contribute an independent semantic value to the overall meaning of the expression but is instead part of a formulaic structure that expresses a particular meaning as a whole. It can thus be argued that this lack of individual semantic contribution makes the presence of the subject *il* optional or redundant, thus contributing to its omission. This semantic distinction across syntactic structures underscores the complex interplay between formal realizations, semantics, and discourse in the possibility of subject omission, even in a non-pro-drop language like French, shaping the distribution of the Null and overt Subjects observed in the analysed data.

6.2. Referential NSs

As mentioned at the beginning of Section 6, out of 3207 referential subjects found within the sub-corpus, only 29 are null. Specifically, three types of referential NSs have been identified, namely, canonical NSs (cf. *infra*) (38a), NSs within repetitions (38b), and NSs referring to an extra-linguistic entity (38c). As can be seen in Table 4, the absolute majority of referential NSs are canonical NSs, while only a few cases of the other two types have been identified.

Table 4. Types of referential NSs.

Type of NS	#	%
Canonical	22	75.86
Repetitions	6	20.69
Extra-linguistic	1	3.45
TOT	29	100

In order to provide a comprehensive analysis of each of the above-mentioned types of NSs, let us now turn to some illustrative examples from the sub-corpus.

(38) a. *Celui sur le pont* là *je ne sais plus*
the one on DET bridge there I not know.1 SG more
comment pro s'appelle
how (he) be.named.3SG
'The one on the bridge there, I don't know his name anymore' (*Auréane*)

b. *Ils attrappent… ouais. pro attrappent des maladies*
they catch.3PL yeah (they) catch.3PL some disease.PL
pas possibles
not possible
'They catch… yeah. They catch some impossible diseases' (*Apéritif*)

c. *pro se cache sous les meubles*
(it) hide.3SG under DET furniture.PL
'It hides under the furniture' (*Montage*)

As can be seen in (38a), the speaker omits the subject of the verb *s'appelle*, which is a third person pronoun coreferent with the entity introduced at the beginning of the sentence (i.e., Celui sur le pont 'the one on the bridge'). This is a 'textbook example' of how NSs occur in consistent pro-drop languages (hence the label 'canonical'), in that the third person NS is linked via Agree to the DP *Celui sur le pont là* which is a specific type of Topic heading a Topic chain (i.e., the A-Topic, cf. Frascarelli 2007). In (38b) the speaker overtly realizes the subject of the verb *attrappent* but then he hesitates and starts again the sentence, omitting the subject that would have been exactly the same. Finally, with the sentence in (38c), the speaker suddenly interrupts his interlocutor, who is talking about something else, referring to a cat that is present in the extralinguistic context, without realizing the relevant overt subject pronoun.

In all these examples, the omitted subjects are linked to a referent that is strongly active in the current discourse context, since they have been introduced as a Topic (38a), uttered mere seconds before (38b), or they are literally in front of the speaker's eyes (38c). Be that as it may, in these cases subject omission can be explained by the presence of an A-Topic, either explicitly introduced as in (38a-b) or silent as in (38c), which establishes an Agree relation with the sentential subject and thus enables speakers to omit it. This result is coherent with what has been proposed in previous studies (cf. Section 4.2), since the A-Topic, which is linked to a canonical NS (cf. 39a), is indeed a left-dislocated prosodically strong constituent. What is more, our analysis suggests that these dislocated constituents may not only be pronouns, but full DP as well (e.g., *Celui sur le pont là*).

In the light of these results and relevant reflections for proposals, it can be now interesting to consider a comparison with another non-pro-drop language. And, this is what we are going to do in the next section, in which a comparison with English will be proposed using the results reported in Cote's (1996) spoken corpora investigation.

6.3. Null Subjects in Non-Pro-Drop Languages: A Comparison Between English and French

Even though different scholars have dealt with subject omission in non-pro-drop languages, focusing on specific contexts of realization, very few works have been systematically dedicated to the possibility and the properties of NSs in a single non-pro-drop language. Among these few studies, Cote's (1996) work on NSs in English represents a precious point of reference for an effective comparison, since relevant results are based on the systematic investigation of a corpus of spoken data.

In particular, Cote used some 10 per cent of the Switchboard Corpus (telephone conversations performed by pairs of native speakers of English adults aged 20–40 on a variety of everyday topics), thus collecting a total of 190 NSs (out of 243 conversations), which have been examined taking into consideration several factors: (a) the form of the subject, (b) the person/number of the subject, (c) the source of the subject (i.e., whether it is referential, deictic, discourse deictic, or expletive), (d) the so-called "centering" transition of the utterance, (e) the turn position of the utterance (i.e., either initial or final), (f) the discourse segment position of the utterance, and (g) the clause type and the sentence type.

Turning to results, it is interesting to notice that the most frequent type of NSs in Cote's corpus of English conversations is not that of expletives, but of referential pronouns (63%). Indeed, null expletives are only 37% of total NSs. However, even if null expletives are not as frequent as in French, they are still much more frequent than their explicit realization even in English, which, in fact, are only 8.9% of total expletive subjects.

Unfortunately, since the author was mainly interested in the discourse-related aspects of subject omission, no specific distinction is provided between expletives. Therefore, no one-to-one comparison can be carried out with respect to our French data. On the other hand, an interesting comparison can be provided between expletive and referential NSs. Indeed, the omission of 1sg and 3sg pronouns appears to be rather frequent in English phone calls, which is contrary to French face-to-face conversations. In particular, null 1sg deictics reach 26% of the total NSs and null 3sg referential pronouns reach 16.6%. In this latter case, it might be interesting to notice that null third person referential subjects were noticeably lacking in animate referents: only 3 out of 30 examples referred to animate entities.

As for discourse-related factors, in Cote (1996) it is reported that 32% of the referential NSs referred to the same entity as did the subject of the previous utterance. Notably, referential NSs in English were mostly used in a continuing function (26.3%); consequently, they served as Given Topics in Topic chains.

As far as their position in the sentence is concerned, Cote's data show that the NS utterances occurred much more frequently in one-utterance turns (38.9% for NSs vs. 6.42% for overt pronominal subjects). Specifically, it seems that NSs in English tend to occur at discourse boundaries (i.e., turn-taking boundaries), while they are rare turn-internally (21.1%). Hence, their function seems to be that of marking a discourse boundary. In this respect, the situation that emerged from our corpus research on Colloquial French is completely different: NSs occur in one-utterance turns only in 14% of total cases and their position is eminently internal, as is shown in Table 5 and the corresponding figure (Figure 1).

Table 5. NSs position (total).

Position	#	%
Initial	39	9.58
Internal	271	66.59
Final	40	9.83
One utterance turn	57	14

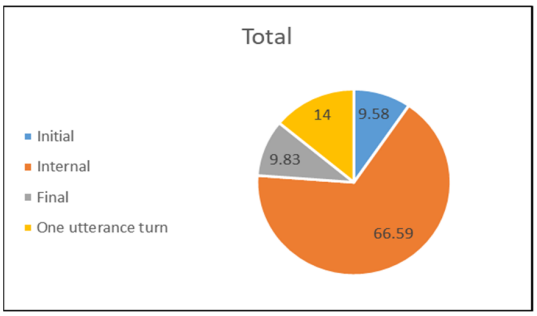


Figure 1. NS position (total).

The dominant preference of NSs for an internal position does not make great distinctions between referential and expletive pronouns, as is shown in Tables 6 and 7 and the corresponding figures (Figures 2 and 3).

Table 6. Position of referential NSs.

Position	#	%
Initial	4	10.26
Internal	24	61.54
Final	3	7.69
One utterance turn	8	20.51

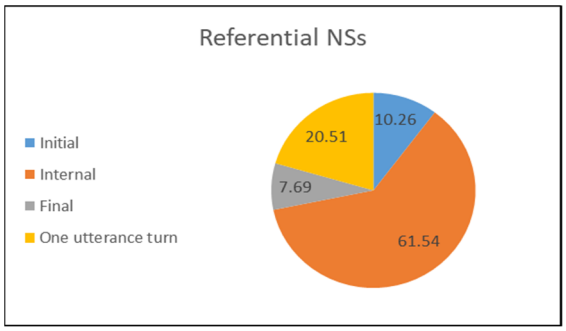


Figure 2. Position of referential NSs.

Table 7. Position of expletive NSs in (*il*) *y a* constructions.

Position	#	%
Initial	30	9.49
Internal	219	69.3
Final	27	8.55
One utterance turn	40	12.66

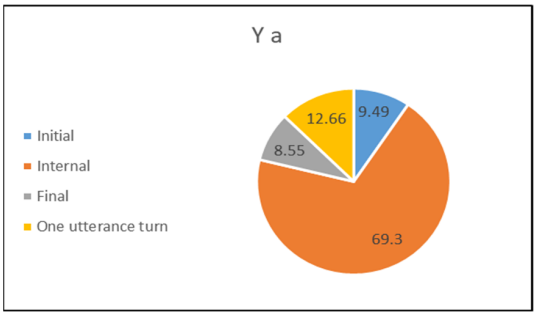


Figure 3. Position of expletive NSs in (il) y a constructions.

Indeed, as can been seen in Table 7 and Figure 3, the percentages attested for the (*il*) *y a construction* (which represents the most frequent realizations for expletive NSs in the corpus examined) show that expletive NSs also prefer an internal position.

In light of the present comparison, we can thus confirm that subject omission is possible in non-pro-drop languages and it is neither sporadic nor occasional. In particular, expletives appear to be the most frequently omitted type of pronouns. Nevertheless, a crucial distinction emerges between French and English, according to which the omission of referential subjects seems to be significantly more frequent in the latter.

Nevertheless, we surmise that this distinction can be attributed to the different context of conversations examined in the relevant corpora: face-to-face (French) vs. phone call conversations (English). Indeed, it is plausible to suppose that a phone conversation between two persons is dedicated and thus concentrates on some specific entity which is taken as the Topic of the relevant discourse and maintained as continuous, somehow "stimulating" its repetition across sentences. On the other hand, Topics can vary during a conversation among friends or be missing while building a piece of furniture.

Of course, these are just feasible assumptions which need to be resumed and confirmed in future comparative studies.

6.4. Coalescence: Corpus Support for a Restricted but Highly Frequent Phenomenon

A side note is reserved in this section to highlight the high frequency of morpho-phonological reduction phenomena occurring in conjunction with specific person–verb inflection associations. We refer to the incorporation of the 1sg subject *je* into the following verb, with the consequent creation of a single lexical form, with related morpho-phonological changes.

Consider the following sentences and the morpho-phonological realization of the subject–verb sequence (IPA transcription in square brackets):

(39)	*Moi*	*j'suis*	**[ʃɥi]**	*pas*	*sûr*	*qu'*	*elle*	*aille*
	pron. 1SG	1SG.CL-be.1SG		NEG	sure	that	pron. 3SGF	have.SUB.3SGF
	trop	*avec*	*les*	*meubles*				
	much	with	the.PL	furniture				
	''I`m not sure it goes too well with the furniture''							(*Apéritif*)
(40)	*Moi*	*j'*	*trouve*	*ça*	*intéressant*	*parce-que*	*moi*	
	pron. 1SG	1SG.CL	find.1SG	it	interesting	because	pron. 1SG	
	j'suis	**[ʃɥi]**	*vraiment*	*nulle*	*en*	*géo*		
	1SG.CL-be.1SG		really	nothing	in	geography		
	''I find it interesting because I'm really bad at geography''							(*Apéritif*)
(41)	*Chomsky*	*j'suis*	**[ʃɥi]**	*un*	*specialiste*	*de*	*Chomsky*	
	Chomsky,	1SG.CL- be.1SG		a	expert	of	Chomsky	
	''Chomsky, I'm an expert of Chomsky''							(*Apéritif*)
(42)	*Julie*	*je sais*	**[ʃɛ]**	*pas*	*qu'*	*elle*	*va*	*prendre*
	Julie,	1SG.CL know.1SG		NEG	what	pron. 3SGF	go.3SG	take
	''As for Julie, I don't know what she's going to take''							(*Apéritif*)
(43)	*Je sais*	**[ʃɛ]**	*pas*	*pourquoi*	*j'*	*ai*	*pas*	*très*	*faim*
	1SG.CL know.1SG		NEG	why	1SG.CL	have.1SG	NEG	much	hunger
	''I don't know why I'm not very hungry''								(*Apéritif*)

As we can see, in these sentences the 1sg subject (weak) pronoun *je* is pronounced as part of the following verb (*suis* 'am' or *sais* 'know'). This fact might be simply ascribed to the fall of the "obsolete" *e* (/ə/ *schwa*), a well-known phenomenon in French, which is often mentioned among scholars and in grammars (also cf. Abeillé and Godard 2021 among others). Nevertheless, based on its specific context of occurrence, we are rather inclined to consider it a particular case of consonant assimilation.[8]

Indeed, corpus analysis shows that this phenomenon does not occur for all the occurrences of the 1sg pronoun and that when it occurs assimilation proceeds in *both* directions. Specifically, the postalveolar fricative [ʒ] of the pronoun *je* determines a change in the place of articulation of the following dental fricative [s], which, in turn, determines the regressive assimilation of the voiceless quality, thus obtaining a postalveolar fricative [ʃ]. On the other hand, the mode of articulation (fricative) remains unchanged.

In such cases, it is therefore appropriate to refer to the notion of 'coalescence' (cf. Zaleska 2020, among others); that is to say, it is a type of assimilation whereby two sounds fuse to become one, and the fused sound shares similar characteristics with the two fused sounds. Some examples in English include 'don't you' -> /dəʊnt ju/ -> [dəʊntʃ u]. In this instance, /t/ and /j/ have fused to [tʃ]. /tʃ/ is a palato-alveolar sound; its palatal feature is derived from /j/ while its alveolar is from /t/. Another English example is 'would you' -> /wʊd ju/ -> [wʊdʒ u]. There are examples in other languages, such as Chumburung where /ìwú ìsá/ -> /ìwúìsá/ becomes [ìwísá]—'three horns'. In this case, /ɪ/ is retained in the coalescence and the rising tone on /u/ appears on the coalesced sound.

Resuming the cases of French illustrated above, corpus analysis shows that these realizations occur almost exclusively with the auxiliary *être* 'to be' and with the verb *savoir* 'know' (occasional occurrences have been found with *je serre* [ʃɛr] 'I squeeze' (1 out of 2) and *je dis* [ʒi] 'I say' (1 out of 1)). Nevertheless, though restricted to these verbs, their frequency is remarkably high: 95% for *je suis* (74 occurrences out of 83), and 73% for *je sais* (45 out of 70). Additional evidence that this phenomenon cannot be (solely, at least) attributed to the fall of the final *schwa* is provided by the presence of a few occurrences of the 2sg pronoun *tu* 'you' and the verb *savoir* 'know' (5 out of 9), obtaining [tɛ] from *tu sais*.

The type of verbs with which this phenomenon occurs seems to support what has been argued in recent works concerning the faster and clearer occurrence of variation phenomena with words of high frequency, as 'to be' and 'to know' undoubtedly are. In particular, in Connine (2004) it is claimed that the representation of auditory form includes explicit representations of the frequently heard variant. Listeners encode surface detail from the speech that they hear and develop lexical representations that match their experience. One consequence of this view is that theoretical accounts of phonological variant processing will be informed by corpus analyses and variant frequency statistics will serve a critical role in theory development for auditory word recognition.

7. Conclusions

This pilot study aimed to investigate the occurrence of NSs in French, trying to provide an answer to two main research questions:

1. What types of subjects (and how often) are indeed allowed to be omitted in Modern Colloquial French?
2. Does the use of NSs in French have patterns similar to other non-pro-drop languages such as English?

Through a corpus study based on three conversations from online corpora of French spoken data, amounting to over 56,000 words, this paper confirms that although French is classified as a non-pro-drop language, there are certain contexts in which NSs do occur. Specifically, *(il) y a* and (*il*) *faut* constructions have been found to be the most likely to occur with a null expletive *il*, while the other verbs were present a lower frequency of NSs. Interestingly, our analysis identified a few instances of NSs occurring with verbs which are not associated with the possibility of subject drop, leading us to suggest that a "list approach" should be abandoned in favour of a more context-oriented approach, aimed at identifying the trigger(s) for the licensing of NSs in a non-pro-drop language like French. In this line of analysis, results show the existence of a crucial distinction between *il* in existential *il y a* constructions, which predicates the existence of the following entity, and *ce* in non-predicational (specificational) sentences, which introduces entities realized as predicates or 'kind events' (Chierchia 1998), since the former is very often omitted while the latter never is.

This difference has been interpreted in light of the fact that *il* is indeed an expletive, and as such devoid of semantic content, whereas *ce* is a referential pronoun. The latter establishes a long-distance Agree relation with a Topic which is thus interpreted as the subject of the SC in a specificational construction. Hence, it cannot be deleted for interpretive requirements at the interfaces. On the other hand, *il* does not have an anaphorical function and is only inserted to meet the EPP requirement.

The relevance of an Agree relation for subject omission also seems to be supported by the few occurrences of referential subject drop attested in the corpus. As proposed in Section 6.2, relevant NSs are all interpreted by virtue of a Topic chain, that is to say, an Agree relations with an A-Topic. This also confirms that in Colloquial French, referential NSs are indeed possible in clauses with a left-dislocated prosodically strong constituent, and as our data indicate, these dislocated constituents may not only be pronouns, but full DP as well.

As for the comparison between French and English, we confirmed that subject omission is possible in both languages, with expletives being the most frequently omitted type of pronouns. Nevertheless, a crucial distinction emerged, according to which the omission of referential subjects seems to be significantly more frequent in English than in French.

Finally, the cases of coalescence found in the corpus (cf. Section 6.4) also support another possible (perhaps concomitant) explanation for the frequent omission of the expletive *il*. As we have shown, the morpho-phonological reduction of the subject only concerns atonic and monosyllabic subjects, hence phonologically defective pronouns. Their transition from clitics to embedding morphemes is a frequent phenomenon in world languages (cf. Givón 1976 and Baker 1998). It is therefore plausible to assume that the third person subject *il*, first reduced to *i-*, incorporating into the locative clitic *y* and giving rise to a long [i:], which, over time, reduced.

These hypotheses clearly need future and more extensive research (also involving phonological analysis) to be supported and eventually corroborated.

Author Contributions: Conceptualization, M.F. and G.C.; methodology, M.F. and G.C.; software, G.C.; validation, M.F. and G.C.; formal analysis, M.F. and G.C.; investigation, M.F. and G.C.; resources, M.F.; data curation, G.C.; writing—original draft preparation, M.F. (Sections 2–4, 6.3 and 6.4), G.C. (Sections 1, 5, 6.1, 6.2 and 7); writing—review and editing, M.F. and G.C.; visualization, M.F. and G.C.; supervision, M.F.; project administration, M.F. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data examined come from the following corpora: (1) The CFPP2000 corpus (http://cfpp2000.univ-paris3.fr/index.html accessed on 1 December 2023): 58 interviews collected after 2000 on the territory of Paris and its neighboring suburbs. A very rich corpus in audio files and texts for downloading, with very detailed descriptions of speakers. (2) The CLAPI corpus (http://clapi.icar.cnrs.fr accessed on 1 December 2023): 194 transcripts documenting oral language in various types of linguistic situations (informal, professional, institutional, and commercial). After examining these two corpora, three transcriptions were selected: "Auréane L'huisser et Pierre-Fabien Benoît" from the CFPP2000 corpus and "Apéritif entre ami(e)s—chat" and "Montage meuble" from CLAPI. A total of two and a half hours of informal conversation was selected, with speakers of medium–high education and aged between 18 and 30–35. Specifically, the first case is a one-and-a-half hour interview with three speakers: two young people in their twenties from the 18th arrondissement of Paris and the interviewer aged around 50. The second case portrays a friendly conversation of about half an hour between four young people, also in their twenties. The third, finally, records the conversation, also of half an hour, of two girls in their twenties struggling with the assembly of an Ikea piece of furniture.

Acknowledgments: We would like to thank the three anonymous reviewers of LANGUAGES for their critical reading of the manuscript and precious suggestions, which helped us improve the final version of this paper. We are grateful to Gianluca Cassetta and Manuela Tomassi for their valuable collaboration in analysing the selected texts for the identification of Null Subjects. We also wish to thank Pierre Larrivée for his noteworthy comments on a first draft of the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

Notes

1 Notice that according to Rizzi's (1982) formulation, if a language has thematic NSs it must also have null expletives. Subsequent analyses, however, have shown that this is only partially true. As Camacho (2013) shows, there are languages that contradict this generalization, like for instance Dominican Spanish, which is an NS language but tends to have overt expletives rather than null ones.

2 The different types of expletives shown below will be resumed in detail in Section 6.

3 Providing support to the fact that there is not a necessary association between the pro-drop quality and rich agreement.

4 The A-Topic connects Reinhart's (1981) aboutness ("what the sentence is about") with the property of being newly introduced or reintroduced to propose a shift in discourse. Assuming with Reinhart that the Common Ground is divided into subsets of propositions that are stored under defining entries (so-called 'file cards'), the A-Topic can be defined as the entry identifying the file card under which the proposition expressed in the sentence is stored. Syntactically, the A-Topic is merged in the highest Topic position in the C-domain (and, from an intonational viewpoint, it is associated with the complex L*+H tone (following the ToBi notation)); i.e., the Topic shift is signaled by a rise in the F0 contour that is aligned with the tonic vowel in its full extension, while the highest point is reached on the post-tonic vowel (cf. Frascarelli 2007).

5 From what we can hear in the recording, the pronoun *il* is not pronounced at all, and thus it qualifies as a case of subject omission. However, as suggested by an anonymous reviewer, in expressions like *s'il vous plaît*, it might be difficult to distinguish a NS from a phonologically reduced one, in which the [l] has been deleted. In any case, due to only two occurrences being found in the corpus, we are unable to make any claim on this issue.

6 In this regard, it may be of interest to note that both necessity and possibility correspond to verbal (hence, predicative) categories, which are realized as formal features in dedicated functional projections in the split-IP cartographic approach (cf. Cinque 1999).

7 In (38), we will use the "trace" notation to indicate positions left empty by movement operations. This is because, although traces have been replaced by "copies" in the Minimalist program, their graphical representation is simpler and more familiar to readers and scholars from the non-generative fields.

8 The literature on assimilation processes is extremely rich (cf. Ohala 1990; Holst and Nolan 1995; Connine 2004; Dilley and Pitt 2007; and Shockey 2008, among many others) and two types of assimilation can be distinguished: progressive and regressive. Regressive assimilation occurs when a following sound has an effect on a preceding one, as in pronouncing 'have to' as 'haf to' for the influence of the voiceless /t/ following /v/. In progressive assimilation the preceding sound has an effect on the following one, as in reverse happens in the plural morpheme /s/ following a voiced consonant (dogs → [dogz]).

References

Abeillé, Anne, and Danièle Godard. 2021. *La Grande Grammaire du Français*. France: Actes Sud.

Baker, Mark. 1998. *Incorporation: A Theory of Grammatical Function Changing*. Chicago: University of Chicago Press.

Barbosa, Pilar. 2011a. Pro-drop and Theories of pro in the Minimalist Program Part 1: Consistent Null Subject Languages and the Pronominal-Agr Hypothesis. *Language and Linguistics Compass* 5: 551–70. [CrossRef]

Barbosa, Pilar. 2011b. Pro-drop and Theories of pro in the Minimalist Program Part 2: Pronoun Deletion Analyses of Null Subjects and Partial, Discourse and Semi pro-drop. *Language and Linguistics Compass* 5: 57–587. [CrossRef]

Belletti, Adriana. 2005. Answering with a cleft: The role of the Null-subject parameter and the vP periphery. In *Contributions to the Thirtieth "Incontro di Grammatica Generativa"*. Edited by Brugè Laura, Giuliana Giusti, Nicola Munaro, Walter Schweikert and Giusepptima Turano. Cafoscarina: Università Ca' Foscari of Venice, pp. 63–82.

Biberauer, Theresa, Anders Holmberg, Ian Roberts, and Michelle Sheehan, eds. 2010. *Parametric Variation: Null Subjects in Minimalist Theory*. Cambridge: Cambridge University Press.

Camacho, José. 2013. *Null Subjects*. Cambridge: Cambridge University Press.

Chierchia, Gennaro. 1998. Reference to kinds across languages. *Natural Language Semantics* 6: 339–405. [CrossRef]

Chomsky, Noam. 1981. *Lectures on Government and Binding*. Dordrecht: Foris.

Chomsky, Noam, and Howard Lasnik. 1977. Filters and Control. *Linguistic Inquiry* 8: 425–504.

Cinque, Guglielmo. 1999. *Adverbs and Functional Heads. A Cross-Linguistic Perspective*. Oxford: Oxford University Press.

Connine, Cynthis M. 2004. It's not what you hear but how often you hear it: On the neglected role of phonological variant frequency in auditory word recognition. *Psychonomic Bulletin & Review* 11: 1084–89.

Cote, Sharon Ann. 1996. Grammatical and Discourse Properties of Null Arguments in English. Ph.D. dissertation, University of Pennsylvania, Philadelphia, PA, USA.

D'Alessandro, Roberta. 2015. Null Subject. In *Contemporary Linguistic Parameters*. Edited by Antonio Fábregas, Jaume Mateu and Michael Putnam. London: Bloomsbury Press, pp. 201–26.
Den Dikken, Marcel. 2006a. Specificational copular sentences and pseudoclefts. In *The Blackwell Companion to Syntax*. Edited by Martin Everaert and Henk van Riemsdijk. Oxford: Blackwell Publishers, vol. IV, chp. 61, pp. 292–409.
Den Dikken, Marcel. 2006b. *Relators and Linkers. The Syntax of Predication, Predicate Inversion, and Copulas*. Cambridge: MIT Press.
Dilley, Laura C., and Mark A. Pitt. 2007. A study of regressive place assimilation in spontaneous speech and its implications for spoken word recognition. *The Journal of the Acoustical Society of America* 122: 2340–53. [CrossRef]
Duguine, Maia. 2013. Null Arguments and Linguistic Variation: A Minimalist Analysis of Prodrop. Ph.D dissertation, Universidad del País Vasco (UPV/EH) and Université de Nantes, Nantes, France.
Frascarelli, Mara. 2007. Subjects, topics and the interpretation of referential pro. An interface approach to the linking of (null) pronouns. *Natural Language and Linguistic Theory* 25: 691–734. [CrossRef]
Frascarelli, Mara. 2010a. Frasi Scisse e 'Small Clauses': Un'Analisi dell'Inglese. *Lingua e Stile* XXXV: 417–46.
Frascarelli, Mara. 2010b. Narrow Focus, Clefting and Predicate Inversion. *Lingua* 120: 2121–214. [CrossRef]
Frascarelli, Mara. 2018. The interpretation of pro in consistent and partial NS languages: A comparative interface analysis. In *Null-Subjects in Generative Grammar. A Synchronic and Diachronic Perspective*. Edited by Federica Cognola and Jan Casalicchio. Oxford and New York: Oxford University Press, pp. 211–39.
Frascarelli, Mara, and Francesca Ramaglia. 2013. Pseudoclefts at the Syntax-Prosody-Discourse Interface. In *The Structure of Clefts*. Edited by Katharina Hartmann and Tonjes Veenstra. Amsterdam and Philadelphia: John Benjamins, pp. 97–138.
Frascarelli, Mara, and Francesca Ramaglia. 2014. The interpretation of clefting (a)symmetries between Italian and German. In *Romance Languages and Linguistic Theory 2012*. Selected Papers from 'Going Romance' Leuven 2012 [RLLT 6]. Edited by Karen Lahousse and Stefania Marzo. Amsterdam and Philadelphia: John Benjamins, pp. 65–89.
Frascarelli, Mara, and Marco Casentini. 2019. The interpretation of null subjects in a radical pro-drop language. Topic chains and discourse semantic requirements in Chinese. *Studies in Chinese Linguistics* 40: 1–45. [CrossRef]
Frascarelli, Mara, and Roland Hinterhölzl. 2007. Types of Topics in German and Italian. In *On Information Structure, Meaning and Form*. Edited by Susanne Winkler and Kerstin Schwabe. Series Editor: Werner Abraham. Amsterdam and Philadelphia: John Benjamins, pp. 87–116.
Freeze, Ray. 1992. Existentials and Other Locatives. *Language* 68: 553–95. [CrossRef]
Gilligan, Gary M. 1987. A Cross-Linguistic Approach to the Pro-Drop Parameter. Ph.D. dissertation, University of Southern California, Los Angeles, CA, USA.
Givón, Thomas. 1976. Topic, pronoun and grammatical agreement. In *Subject and Topic*. Edited by Charles N. Li. New York: Academic Press, pp. 149–88.
Grimshaw, Jane, and Vieri Samek-Lodovici. 1998. Optimal subjects and subject universals. In *Is the Best Good Enough? Optimality and Competition in Syntax*. Edited by Pilar Barbosa, Danny Fox, Paul Hagstrom, Martha McGinnis and David Pesetsky. Cambridge: MIT Press, pp. 193–219.
Haegeman, Liliane. 2000. Adult null subjects in non pro-drop languages. In *The Acquisition of Syntax: Studies in Comparative Developmental Linguistics*. Edited by Marc-Ariel Friedemann and Luigi Rizzi. Harlow: Longman, pp. 129–69.
Holmberg, Anders. 2005. Is there a little pro? Evidence from Finnish. *Linguistic Inquiry* 36: 535–64. [CrossRef]
Holmberg, Anders. 2010. Parameters in minimalist theory: The case of Scandinavian. *Theoretical Linguistics* 36: 1–48. [CrossRef]
Holmberg, Anders, Aarti Nayudu, and Michelle Sheehan. 2009. Three partial null-subject languages: A comparison of Brazilian Portuguese, Finnish and Marathi. *Studia Linguistica* 63: 59–97. [CrossRef]
Holmberg, Anders, and Urpo Nikanne. 2002. Expletives, subjects, and topics in Finnish. In *Subjects, expletives, and the EPP*. Edited by Peter Svenonius. Oxford: Oxford University Press, pp. 71–106.
Holst, Tara, and Francis Nolan. 1995. The influence of syntactic structure on [s] to [∫] assimilation. In *Phonology and Phonetic Evidence: Papers in Laboratory Phonology IV. Papers in Laboratory Phonology*. Edited by Bruce Connell and Amalia Arvaniti. Cambridge: Cambridge University Press, pp. 315–33.
Huang, C. T. James. 1984. On the Distribution and Reference of Empty Pronouns. *Linguistic Inquiry* 15: 531–74.
Ingham, Richard. 2018. Topic, Focus and null subjects in Old French. *Canadian Journal of Linguistics/Revue Canadienne de Linguistique* 63: 242–63. [CrossRef]
Jaeggli, Osvaldo, and Ken Safir, eds. 1989. *The Null Subject Parameter*. Dordrecht: Kluwer.
Kayne, Richard. 1980. Extensions of Binding and Case-Marking. *Linguistic Inquiry* 11: 75–96.
Krifka, Manfred. 2006. Basic notions of information structure. In *Interdisciplinary Studies on Information Structure (ISIS) Working Papers of the SFB632, 6*. Edited by Caroline Féry, Gilbert Fanselow and Manfred Krifka. Potsdam: Universitätverlag Potsdam, pp. 13–56.
La Fauci, Nunzio, and Michele Loporcaro. 1997. Outline of a theory of existentials on evidence from Romance. *Studi Italiani di Linguistica Teorica e Applicata* 26: 5–55.
Langacker, Ronald W. 1985. Observations and speculations on subjectivity. In *Iconicity in Syntax*. Edited by John Haiman. Amsterdam and Philadelphia: John Benjamins, pp. 109–50.
Mayol, Laia. 2010. Contrastive pronouns in null-subject Romance languages. *Lingua* 120: 2497–514. [CrossRef]
Modesto, Marcelo. 2008. Topic prominence and null subjects. In *The Limits of Syntactic Variation*. Edited by Theresa Biberauer. Amsterdam and Philadelphia: John Benjamins, pp. 375–410.

Napoli, Donna Jo. 1982. Initial Material Deletion in English. *Glossa* 16: 85–111.
Neeleman, Ad, and Krista Szendröi. 2007. Radical pro-drop and the morphology of pronouns. *Linguistic Inquiry* 38: 671–714. [CrossRef]
Ohala, John J. 1990. The phonetics and phonology of aspects of assimilation. *Papers in Laboratory Phonology* 1: 258–75.
Perlmutter, David. 1971. *Deep and Surface Constraints in Syntax*. New York: Holt, Rinehart and Winston.
Poletto, Cecilia, and Jean-Yves Pollock. 2004. On the left periphery of some Romance wh-questions. In *The Structure of CP and IP. The Cartography of Syntactic Structures, Vol. 2*. Edited by Luigi Rizzi. Oxford: Oxford University Press, pp. 251–96.
Ramaglia, Francesca, and Mara Frascarelli. 2019. The (information) structure of existentials. In *Linguistic Variation: Structure and Interpretation*. Edited by Ludovico Franco and Paolo Lorusso. Berlin and Boston: Mouton De Gruyter, pp. 561–98.
Reinhart, Tania. 1981. Pragmatics and Linguistics: An analysis of Sentence Topics. *Philosophica* 27: 53–94. [CrossRef]
Rinke, Esther. 2003. On the licensing of null subjects in Old French. In *Syntactic structures and morphological information*. Edited by Uwe Junghanns and Luka Szucsich. Berlin: Mouton De Gruyter, pp. 101–27.
Rinke, Ester. 2007. *Syntaktische Variation aus Synchronischer und Diachronischer Perspektive*. Frankfurt am Main: Vervuert Verlagsges.
Rizzi, Luigi. 1982. *Issues in Italian Syntax*. Dordrecht: Foris.
Roberts, Ian, ed. 1993. *Verbs and Diachronic Syntax*. Dordrecht: Kluwer.
Saito, Mamoru. 2007. Notes on East Asian Argument Ellipsis. Nanzan University and University of Connecticut. *Unpublished manuscript*.
Shockey, Linda. 2008. *Sound Patterns of Spoken English*. Hoboken: John Wiley & Sons.
Stowell, Tim A. 1981. Origins of Phrase Structure. Ph.D. dissertation, MIT, Cambridge, MA, USA.
Taraldsen, Knut Tarald. 1980. *On the Nominative Island Constraint, Vacuous Application and the That-Trace Filter*. Bloomington: Indiana University Linguistics Club.
Tomioka, Satoshi. 2003. The semantics of Japanese null pronouns and its cross linguistic implications. In *The Interfaces: Deriving and Interpreting Omitted Structures*. Edited by Kerstin Schwabe and Susanne Winkler. Amsterdam and Philadelphia: John Benjamins, pp. 321–39.
Torres Cacoullos, Rena, and Catherine E. Travis. 2014. Prosody, priming and particular constructions: The patterning of English first-person singular subject expression in conversation. *Journal of Pragmatics* 63: 19–34. [CrossRef]
Tsao, Feng-Fu. 1977. A Functional Study of Topic in Chinese: The First Step Toward Discourse Analysis. Ph.D. dissertation, USC, Los Angeles, CA, USA.
Vance, Barbara. 1997. *Syntactic Change in Medieval French: Verb-Second and Null Subjects*. Dordrecht: Kluwer.
Wagner, Susanne. 2012. Null Subjects in English: Variable Rules, variable Language? Ph.D. dissertation, Institut für Anglistik und Amerikanistik, Chemnitz University of Technology, Chemnitz, Germany.
Ylinärä, Elina, and Mara Frascarelli. 2021. Subjects, Topics and the ACC Case alternation in Finnish. A discourse-related "split" account for the Subject function. *Lingua* 255: 1–28. [CrossRef]
Zaleska, Joanna. 2020. Coalescence as autosegmental spreading and delinking. *Phonology* 37: 697–735. [CrossRef]
Zimmermann, Michael. 2009. On the evolution of expletive subject pronouns in Old French. In *Proceedings of the Workshop on "Null-Subjects, Expletives and Locatives in Romance"11*. Edited by Georg A. Kaiser and Eva-Maria Remberger. Konstanz: Konstanz Universität, Fachbereich Sprachwissenschaft, pp. 63–92.
Zimmermann, Michael. 2014. *Expletive and Referential Subject Pronouns in Medieval French*. Berlin: Mouton de Gruyter.
Zimmermann, Michael. 2018. Null subjects, expletives, and the status of Medieval French. In *Null Subjects in Generative Grammar: A Synchronic and Diachronic Perspective*. Edited by Federica Cognola and Jan Casalicchio. Oxford: Oxford Academic, pp. 70–93.
Zimmermann, Michael, and Georg A. Kaiser. 2014. On expletive subject pronoun drop in Colloquial French. *Journal of French Language Studies* 24: 107–26. [CrossRef]

languages

MDPI

Article

Complementizer Agreement and the Licensing of DPs: An Account in Terms of Referential Anchoring

Roland Hinterhölzl

Department of Linguistics and Cross-Cultural Studies, Università Ca' Foscari, 30123 Venezia, Italy; rolandh@unive.it

Abstract: In this paper, I argue that the phenomenon of complementizer agreement in West Germanic and the distribution of DPs in German can be given a common explanation in terms of an approach in which context values are not freely assigned via an interpretive function operation, as is assumed in standard accounts of formal semantics, but rather, they become accessible in a specific functional head in the C-domain.

Keywords: complementizer agreement; clitic movement; scrambling; weak and strong definite determiners; anchoring to the discourse; context values

check for updates

Citation: Hinterhölzl, Roland. 2024. Complementizer Agreement and the Licensing of DPs: An Account in Terms of Referential Anchoring. *Languages* 9: 49. https://doi.org/10.3390/languages9020049

Academic Editor: Ana Ojea

Received: 23 November 2023
Revised: 11 January 2024
Accepted: 13 January 2024
Published: 29 January 2024

1. Introduction

In this paper, I address the well-known issue that DP-arguments, dependent on their interpretation, occupy different domains in the clause in many languages. For instance, definite DPs in German occupy higher positions in the middle field than indefinite ones (cf. Kratzer 1989; Diesing 1992). It is assumed that indefinite DPs can be licensed in the V-domain, while definite ones move higher and are presumably licensed in the T-domain.[1]

In addition, discourse anaphoric DPs and pronominal DPs move to even higher positions in the clause in German and can be argued to be licensed in the C-domain. Furthermore, there is the phenomenon of complementizer agreement in many Germanic dialects that is still lacking an intrinsic motivation. It is generally assumed that a C-head enters in an agree relation with a finite verb (in T). However, this remains a stipulation. I will argue in this paper that complementizer agreement is a reflex of an intrinsic licensing relation between Fin^0 and the temporal argument structure of the verb, on the one hand, and referential DPs, crucially including subject pronouns, on the other hand.

We may ask what the reason for the movement of definite DPs (and pronouns) is, given that the standard semantic account of the interpretation of referential expressions is in terms of assignment functions that assign a referential index to a DP from a context set. If this were correct, DP-licensing would be possible in any position of the clause.

I note that case or agreement with a finite verb cannot be taken to motivate these movement operations either, since indefinite DPs also have case and can, as subjects in a presumed vP internal position, enter into an agree relation with T (or AgrS, depending on one's favorite theoretical assumptions).

I will argue that the reason behind this distribution is that context values are not freely assigned but become accessible only in the C-domain that, as advocated by Rizzi (1997), serves to connect the proposition with the context.

In particular, I will assume that context values, that is, values for established discourse referents, on the one hand, and for the reference situation that is crucial for the temporal anchoring of the clause, on the other hand, are accessible in Fin^0.

Furthermore, I will argue that the more fine-grained distribution of indefinite DPs (in the V-domain), weak (definite) DPs (in the T-domain), and strong (definite) DPs (in the C-domain) follows from the presence or lack of presuppositional requirements that

the different determiners impose on the individual and the situation argument of the nominal head.

2. DP Types and Their Role in the Discourse

Let us start out with a discussion of the different discourse roles of definite and indefinite DPs. It was proposed first by Irene Heim (1982) in her famous familiarity condition that indefinite DPs serve to introduce new discourse referents, while definite DPs serve to pick up (i.e., refer back to) referents that have been introduced in *the* previous discourse. While this is certainly correct for typical cases of the use of definite DPs, it was soon after noticed that there are uses of definite DPs that also introduce a new discourse referent.

In fact, Donnellan (1966) had pointed out that one needs to distinguish between the referential use of a definite description, which obeys Heim's familiarity condition, and the attributive use of a definite description, in which it is essential that there be a uniquely identifiable referent (typically not yet given in the existing discourse) in the situation that is at issue, as in (1). The referent of *the winner* in (1) is not known at the point of the utterance, but it is uniquely identifiable as soon as the race has been decided.

1. Tomorrow there will be a 100 m run in Vienna.
   The winner will receive a Porsche!

Thus, there are two different conditions, the familiarity condition and the uniqueness condition, that both seem to be relevant for the use of the definite determiner. The dispute, in fact, goes back to Russell (1905) and Frege (1892). While the former proposed that the definite determiner purports an assertion as to the existence of a unique individual that fulfills the nominal predicate and is, thus, apt to account for the attributive use of a definite description, Frege (1892) held that a definite description imposes a presupposition that there exists an individual that fulfills the nominal predicate.

As far as (1) is concerned, Russell's treatment would foresee the complex assertion in (2), while Frege would argue that the presupposition of the definite description in (1) is fulfilled, since when there is a race there is a unique winner, and (1) amounts to the assertion that this individual (the one that fulfills the presupposition) will receive a Porsche.

2. ιx winner (x) & will-receive-a-Porsche (x)

I will assume in this paper that both are right, and I will distinguish between the weak and the strong definite determiner in Germanic. It has long been noted that several Germanic languages/dialects have two full article paradigms (cf. Heinrichs (1954) for Rhineland dialects, Scheutz (1988) and Schwager (2007) for Bavarian, and Ebert (1971) for the Frisian dialect of Fering). In standard German, the distinction becomes apparent in certain preposition–article combinations, as is illustrated in (3):

3.	a.	Hans	ging	in-s	Haus.	(D-weak)
		John	went	into the	house.	
	b.	Hans	ging	in das	Haus.	(D-strong)
		John	went	into the	house.	

While (3a) can be uttered out of the blue, (3b) is only possible if the relevant house has already been mentioned in the previous discourse. Thus, it appears that we have to deal with two different types of definite determiners that also differ in their semantics. Schwarz (2012) argued that one should not strive for a unified theory of the semantics of the definite determiner in Germanic since the weak definite determiner is subject to a uniqueness requirement, while such a requirement is apparently irrelevant for the strong definite determiner. Readers are referred to Schwarz (2012) for further details.

Following Frege (1892), I propose that the uniqueness condition is also relevant for the strong determiner but only in the definition of the presupposition that serves to discriminate the antecedent of the strong definite DP in the discourse.

Furthermore, I will make the following proposal to solve the question about the definition of the situation in which the uniqueness condition imposed by a definite determiner

is supposed to hold. While in standard treatments of definite descriptions, as in Schwarz (2012) and others, it has been assumed that a situation pronoun is introduced by a definite determiner, and hence, it is absent in indefinite DPs, I propose that this situation argument is introduced already by the nominal head, and hence, it is also available in indefinite DPs.

In other words, every nominal referent is individuated with respect to a situation. However, definite DPs, both weak and strong ones, impose a presupposition on the identifiability of this situation argument in the common ground (CG), while indefinite DPs come without any presupposition on this argument. This means that a definite determiner indicates that the situation argument of a nominal is, in some sense, given, while an indefinite determiner indicates via an implicature that the situation argument of a nominal is not given in this sense. Thus, both weak and strong definite determiners share the property of imposing the same usage condition on a definite description that is distinct from the usage condition of an indefinite description.

The difference between weak and strong definite determiners only concerns the individual argument of a nominal. With a weak determiner, it is asserted that there is a unique individual in a situation given in the CG, while with a strong determiner, it is presupposed that there is a unique individual identifiable in the CG, which satisfies the nominal predicate in the situation given in the CG, as illustrated in (4). In (4), conditions that operate as presuppositions are put before the dot of the lambda-operator and are underlined, while conditions that are asserted appear after the dot of the lambda-operator.

4. a. $[\![D]\!] = \lambda P\ \underline{\exists s\ s \text{ in CG}}\ .\ \iota x\ P\ (x, s)$ (weak definite determiner)
   b. $[\![D]\!] = \lambda P\ \underline{\exists s\ s \text{ in CG} \ \&\ \iota x \text{ in CG} \ \&\ P\ (x, s)}.\ x$ (strong definite determiner)

This means that if an indefinite DP is merged in the vP, its situation argument, being without any presupposition, can be identified with the event argument of the verb, while a definite DP cannot be licensed in the vP since its presupposition requires that its situation argument is identified with a situation that is already in the CG. One specific situation that is already in the CG and becomes available in the T-domain is the reference situation that plays a crucial role in the temporal interpretation of the predicate, as we will see in the next section.

As far as pronouns are concerned, I propose that they also have a nominal core, namely, the abstract nominal predicate *participant* that relates an individual argument and a situation argument, as is illustrated for a personal pronoun in (5) and for a deictic pronoun in (6). The features in D are interpreted as presuppositional conditions on an individual referent that serve to discriminate the discourse antecedent in the context. The situation argument of personal pronouns is identified with the reference situation (s_R) (see below), while the situation argument of deictic pronouns is identified with the utterance situation (s_U). These feature specifications require pronouns to be licensed in the T-domain for their situation argument and to have access to the C-domain for their individual argument.

5. he = $[_{DP}$ D $[_{nP}$ participant (x, s_R) $]]$
   3SG, male
6. we = $[_{DP}$ D $[_{nP}$ participant (x, s_U) $]]$
   1PL

Summing up this section, I note that Diesing's account can explain why indefinite DPs need to be licensed in the V-domain, but it fails to account for why definite DPs must move out of the V-domain. The present account, on the other hand, explains why indefinite DPs can be licensed in the V-domain and why definite DPs must move out of the V-domain to be licensed in the T- or C-domain. In particular, I will argue that weak definites are licensed in the T-domain while strong (referential) definites are licensed in the C-domain.

3. The Reference Situation and the Anchoring of a Thetic Judgment

It is standardly assumed that a predicate is anchored via tense (and mood) to a context. Interested readers are referred to Zagona (2013) and the references cited therein for further background on this issue. In particular, it is assumed that tense (in a matrix clause) locates

the verbal event with respect to the utterance time. Thus in (7), a speaker asserts that there was an event of visiting in the past (before the utterance time) in which an individual named John functioned as the agent (the visitor) and his mother functioned as the theme (the visitee of the event).

7. a. John visited his mother.
   b. ∃e visiting(e) & past (e) & agent (e, John) & theme (e, his mother)

However, this simple linking approach is insufficient when we look at examples embedded in a discourse, as illustrated in (8). Anaphorically linking *she* to *his mother* in (8a), and simply anchoring the event e_2 to the utterance time, the meaning of (8b) would be compatible with this event preceding, following, or overlapping with e_1 as long as both events precede the utterance time (s), as illustrated in (8c). This rendition is incomplete, since speakers typically interpret (8b) as a claim about John's mother being sick at the time of his visit. Moreover, the adverbials in (8d) are also interpreted with respect to John's visit.

8. a. John visited his mother. (e_1)
   b. She was sick. (e_2)
   c. $e_1 < e_2 < s$, $e_2 < e_1 < s$, or e_1 o $e_2 < s$
   d. She was sick one week before/earlier.

The problem can be solved by introducing a reference situation. Here, I am following Reichenbach (1947), according to whom tense establishes a link between a speech time and a reference time, as illustrated in (9). The event argument of the verb is then taken to be situated with respect to the reference time by aspect, as is illustrated in (10).

9. The meaning of tense according to Reichenbach (1947):
   a. past: = r < s
   b. present: = r ⊆ s
10. The meaning of aspect according to Reichenbach (1947):
   a. perfect: = e < r
   b. imperfect: = e ⊂ r

As is indicated in (8), the event in (8a) serves as discourse antecedent for the interpretation of tense in the clause in (8b) and for the temporal adverbials in (8d). Hence, I will propose that tense is not a predicate of a point of time or an interval, as is standard since the seminal work on tense by Stowell (1995), but rather, that tense relates two situations (an utterance situation and a reference situation). Hence the temporal interpretation itself is secondary and derived from a relation between situations making use of the running time of a situation (τ), as illustrated in (11).

11. Situation-based account of tense (SAT):
   - Tense is a predicate that relates situation arguments.
   - past (s_1, s_2) = s_1 precedes s_2 = : τ (s_1) < τ (s_2)

Normally, a predicate is anchored to a context by a definite subject, as in (12a). The result is a categorical judgment about a particular individual (or a particular set of individuals). Alternatively, a predicate can be anchored to a context via a reference situation, as in (12b). The result is a thetic judgment about a particular situation. It was argued in Hinterhölzl (2024) that *e*s in German is not an expletive element but (being inserted in [Spec, TP]) instead binds a reference situation argument of tense. An anonymous reviewer asked why *es* in (12b) cannot be taken to be inserted in the C-domain since it serves to distinguish between a judgment and a question, properties that are taken to be defined in the C-domain. I assume that sentence types (declarative or interrogative moods) are defined in ForceP and that FinP, together with TP, serves to referentially anchor a proposition.

12. a. Hubert Haider spricht.
       Hubert Haider speaks.
   b. Es spricht Hubert Haider.
       It speaks Hubert Haider.

More importantly, *es* becomes obligatory if a subject semantically cannot serve as an anchor because it is indefinite as in existential constructions, as in (13a), or if a predicate does not have an argument of itself, as with weather verbs, in which case the verb is

predicated of the reference situation argument of tense, as illustrated in (13b–d). In (13d), s_U refers to the utterance situation. Thus, both (13a) and (13b) constitute thetic judgments about a specific situation.

13. a. Es gab einen Aufruhr.
    It gave an uprise (there was an uprise).
    b. Es regnete.
    It rained.
    c. s_1 (that is identified with the reference situation) $\in$ {s | rains in s}
    d. $\lambda P(s,s_U).\ \iota s\ P(s,s_U)$ (meaning of es)

In English sentences with an indefinite subject, the adverbial *there* is inserted in Spec, TP. Also, here, I argue that *there* is not an expletive element but instead serves semantically as an alternative anchor in the clause, as illustrated in (14).

14. a. John visited his mother.
    b. There was a child crying in the garden.
    c. I went to the local bar last night. Into the room walked a man with a green hat

In the present account, *there is* a function that maps the reference situation onto its location and refers back to the situation of John's visit that provides the situation with respect to which the predicate 'was a child crying in the garden' is temporally and locally evaluated. In a similar vein, a PP, by denoting the resultant location of a predicate expressing a change in state (or location), can serve as a subject/anchor in the case of locative inversion, as illustrated in (14c). In (14c), *into the room* refers to the room in the previously mentioned bar situation from the previous night. Thus, both (14b) and (14c) qualify as thetic judgments.

As already argued for by Milsark (1974), *there* cannot be treated as an expletive element that is replaced at LF by a real subject, and this is shown by the observation that the subjects in (15) have the following different interpretations: in (15a), the subject has a weak cardinal interpretation, and in (15b), the subject has a strong proportional interpretation. Interested readers are referred to Hinterhölzl (2019) and the references given therein for a thorough account of the syntax and semantics of the expletive construction in English.

15. a. There were not many people in the room.
    b. Many people were not in the room.

It is interesting to note that strong quantifiers can anchor a predicate to a context but need not do so, as illustrated in (16). In (16), taken from Schwarz (2012) and also discussed in Hinterhölzl (2019), the subject *most senators* has a strong proportional reading, but the sentence appears to characterize the political situation in 2004, constituting a thetic judgment.

16. What was the political situation in congress in 2004?
    Most senators were Republicans.

Arguably, we have a case in which an attributively used DP is evaluated with respect to a given situation, hence the strong interpretation. I propose that the subject in this case is licensed in the T-domain by identifying its situation argument with the reference situation of tense. The speaker in (16) does not make reference to a specific group of senators, but rather, is simply stating that a majority of the senators at that time were Republicans.

In conclusion, we propose that indefinite DPs and weak quantifiers are interpreted in the V-domain while attributively used definite DPs and strong quantifiers (if not discourse-anaphoric) are interpreted in the T-domain. Furthermore, I propose that referentially used definite DPs and anaphoric strong quantifiers require access to the C-domain to be fully licensed. In particular, I propose that discourse-anaphoric DPs must enter into a licensing relation with the head Fin^0.

In the following section, I will discuss the distribution and the licensing of subjects in Cimbrian. These data will provide an interesting parallel to the patterns found in complementizer agreement in languages/dialects that allow for double agreement.

4. Subjects in Cimbrian: A Case Study

Let us take a look at the distribution of subjects in Cimbrian, a German dialect spoken in the village of Luserna, Trentino. Field work was carried out by Federica Cognola, and the data were published in Cognola and Hinterhölzl (2020). Interested readers are referred to this article for a more complete picture of the complex interaction between V2, question formation, and the licensing of subjects in this variety. As is illustrated in (17), there is a complementary distribution between preverbal subjects and the presence of a subject pronoun, or *da* ('here, there'), cliticized on to the verb. In (17), *da* is spelled out as *-ta* when cliticized on the verb. The subject in (17f) is unmarked with respect to its information structural role (i.e., it can be new or given information), as follows:

17. a. Bas hatt-ar$_j$ herta gekoaft dar Luca$_j$?
What has he always bought the Luca?
b. Bas hat-ta herta gekoaft dar Luca?
What has_*da* always bought Luca?
c. *Bas hat herta gekoaft dar Luca?
What has always bought the Luca?
"What has always Luca bought?"
d *Dar Luca hatt-ar$_j$ herta gekhoaft in libar.
The Luca has he always bought a book.
e. *Dar Luca hat-ta herta gekhoaft in libar.
The Luca has *da* always bought a book.
f. Dar Luca hat herta gekhoaft in libar.
The Luca has always bought a book.
"Luca has always bought a book."

This characterization can also be found in various publications about the role of *da* in Cimbrian (Bidese and Tomaselli 2005 and subsequent work; Kolmer 2005; Grewendorf and Poletto 2015), where *da* and a subject clitic are ruled out in all cases in which a subject precedes a finite verb, and *da* or a subject clitic are obligatory in all cases in which a subject follows a finite verb.

As I will argue below, movement of a constituent into the preverbal domain interferes with the licensing of subjects. This is due to the V2 nature of the language where [Spec, FinP] constitutes a bottleneck for movement into the C-domain.[2] In other words, the wh phrase *bas* in (17) has to pass through [Spec, FinP] to reach [Spec,ForceP] to license the speech act of a question. Subjects in this case can be licensed in a lower position when they are doubled by *da* or by a subject clitic pronoun, as follows: if doubled by *da,* the subject is focused (new information or a contrastive focus), and if doubled by a pronoun, the subject is a topic. The main stress in (18a) falls on the sentence final subject while the sentence final subject in (18b) is optional and unstressed.[3]

18. a. Haüt iz = ta khent dar nono
Today is *da* arrived the grandfather.
b. Haüt izz = ar$_j$ khent (dar nono$_j$)
today is = he arrived (the grandfather)
c. *Haüt iz khent dar nono
Today is arrived the grandfather.
"The grandfather arrived today."

We may wonder what the roles of clitics and *da* are in the licensing process of the lower subject in (18ab). Let us first discuss what is said about *da* in the literature. *Da* is only homophonous with the locative *da* ('here') (see Grewendorf and Poletto 2015, p. 402; Kolmer (2005); and Bidese and Tomaselli (2018 and previous work)). As illustrated in (19), *da* cliticized onto a main verb can occur with an instance of the locative *da*.

19. Bas hat-ta gatont a khin *da*? (Grewendorf and Poletto 2015, p. 402)
What has *da* done a boy there?
"What has a boy done there?"

Da differs from the English 'there', realizing [Spec,TP], since *da* is compatible with definite and indefinite NPs and nothing can intervene between *da* and a finite verb in main clauses and between *da* and the complementizer *bo* (cf. Bidese et al. 2012).

Furthermore, it is uncontroversial that *da* is hosted in the lower portion of CP, i.e., FinP (see Rizzi 1997), and that its position with respect to a finite verb is fed by V-to-C movement (see Bidese and Tomaselli 2005 and subsequent work and Grewendorf and Poletto 2015 for an analysis of Cimbrian as a V2 language).

The idea that I would like to develop in the following section is that clitics and *da*, by undergoing head movement to Fin^0, serve to license the subject when the latter is unable to undergo movement to [Spec, FinP] for syntactic (i.e., another element moves through [Spec, FinP]) or semantic reasons (i.e., the subject is indefinite).

5. The Special Role of Clitic Pronouns and *da* in Anchoring the Utterance to the Context

In this section, I will argue that clitics and *da* serve to referentially anchor a subject to a context. In particular, I propose that definite DPs, in contradistinction to indefinite DPs, have an extra layer, which is an additional functional head that licenses a correlate DP in its specifier, as in (20).

Indefinite and weak DPs lack the respective layer and, thus, must combine with an adverbial alternative anchor, as occurs with the English existential 'there' construction.

20. [$_{DP}$ [$_{DP}$ *da*/cl] [D^0 [$_{NP}$ N]]

Referential subjects can always anchor a predicate and obtain access to the value of their discourse antecedent if they move into a pre-finite position, that is, into [Spec, FinP] in Cimbrian, as illustrated in (21). Furthermore, I propose that if a referential DP is moved into [Spec, FinP], no correlate is generated in [Spec, DP] for reasons of economy, explaining the ungrammaticality of (21b).

21. a. Dar Mario hat gekhoaft in liber.
    The Mario has bought the book.
    b. *Dar Mario$_j$ hat-ta/hat-ar$_j$ gekhoaft in liber.
    The Mario has *da*/has he bought the book.
    "Mario bought the book."

Non-definite subjects can anchor an utterance to a context if the DP has a strong interpretation (QN or WhN), as illustrated in (22). Also, in this case, a correlate clitic or *da* are excluded for reasons of economy.

22. a. Belz khinn hatt bokhennt soin tatta?
    Which child has met his father?
    b. *Belz khinn hatt-ta bokhennt soin tatta?
    Which child has *da* met his father?
    "Which kind met his father?"

With non-definite subjects without any NPs overtly realized (e.g., bare QPs and simple wh elements), the following two cases must be distinguished: (a) non-subject questions trigger always the presence of a clitic or *da* depending on whether the lower subject is to be interpreted as a topic or as a focus, as illustrated with a focused subject again in (23).

23. Bas hat-ta herta gekoaft dar Luca?
    What has *da* always bought the Luca?

Here, the idea is that since the subject is blocked by wh movement to move into [Spec, FinP], the correlate sub-extracts from the subject DP, which remains in a lower position and undergoes head movement to Fin^0 to connect the subject with a specific referential value from the context set, where the features of the clitic serve to discriminate the relevant discourse antecedent.

With subject questions, *da* is optional depending on the interpretation of the subject, as illustrated in (24) and (25). Since the wh subject is moved through [Spec,Fin], it alone can anchor the utterance if it has a definite interpretation, as in (24). If it has an indefinite interpretation, the adverbial *da* must be inserted in [Spec,TP] and undergo head movement to Fin^0.

The data in (24) and (25) are taken from an empirical investigation, i.e., interviews with native speakers of Cimbrian in Luserna carried out by Federica Cognola and reported in Cognola and Hinterhölzl (2020). Interested readers are referred to this paper for the details of this study. In (24) and (25), the native speaker's judgments are given by an evaluation on the Likert scale between zero (ungrammatical) and five (fully grammatical). Since the context in (24) triggers an indefinite interpretation on the wh subject, only the version with

da is possible. Since the context in (25) triggers a definite interpretation on the wh subject (who of us), the version without *da* is fully grammatical.

24. Context: You are watching TV and hear the telephone ringing. You ask the following:
    a. Ber riüft-ta o? → 4,8/5
       Who calls *da* up?
    b. Ber riüft o? → 2/5
       Who calls up?
       "Who is calling?"
25. Context: You and your friends have to book a room for the weekend. You do not know who is supposed to call the hotel. You ask the following:
    a. Ber riüft-ta o? → 2,5/5
       Who phone *da* up?
    b. Ber riüft o? → 4,8/5
       Who phone up?
       "Who of us is going to make the call?"

An anonymous reviewer asked why it is that only subjects interact with clitic pronouns and *da* in Cimbrian, while it must be assumed that all referential DPs (of objects and prepositional objects) must have access to Fin^0. The latter assumption is correct. I propose that Fin^0 enters into an agree relation with all the referential constituents contained in TP and values them, but it will only attract the referential subject since it constitutes the highest argument in the structure. Given that it is the movement of the subject into [Spec,FinP] that interferes with the wh movement via the bottleneck effect, it is subjects that interact with the presence/absence of clitic pronouns and *da* in Cimbrian. When a subject is non-referential (i.e., indefinite), it will remain in a lower position and the sentence will be anchored via the reference situation argument of tense as a thetic judgment. Here, I will leave aside the issue of anchoring referential adverbials.

To sum up what we have found so far, referential DPs cannot be interpreted within vP without any additional operation that connects them with the C-domain (see (17b) above, repeated here as (23)). I have proposed that referential DPs have an extra layer. A definite strong DP, when unable to move to [Spec, FinP] for syntactic reasons, is licensed by movement of a correlate adnominal *da* or a pronominal correlate, and the choice is language- or function-specific (topic vs. focus). Note in particular that some languages also allow clitics with focused constituents, such as Spanish and Romanian. However, I cannot address this issue in any detail in this paper. In the following section, I will address the question of what happens in cases where the subject is not anchored via an anaphoric link.

6. The Role of Frame Adverbials

Frame adverbials play a special role in the anchoring process of statements. I first note that IP-related temporal (and locative) adverbs express a relation between the reference situation and the event time/location, as we have already seen in (8d) above.

Frame adverbials crucially have a different interpretation. They shift or restrict the reference situation itself, as illustrated in (26) and (27). While speaker A in (26) talks about Christmas in the past, speaker B shifts the reference point with the expression 'in not many years' to a future reference situation. Likewise, in (28), the adverbial 'with no job' restricts the set of people that would be happy, giving rise to a strong proportional reading of the quantifier 'few'.

26. A: Last year, Christmas was fun. We had 5 days of free holidays.
    B: In not many years, Christmas will fall on a Wednesday again.
27. With no job, few people would be happy.

An initial check of Cimbrian data indicates that generic statements in Cimbrian always appear without *da* or a subject clitic, but this was not investigated in detail by Cognola and Hinterhölzl (2020). If this observation is verified on a larger dataset, it would imply that the subject is anchored in a different way in these cases.

Here, I will limit myself by motivating this claim with English data. As illustrated in (28), a frame adverbial such as 'in Australia' restricts the set of swans to Australian swans.

I propose that in this case, the situation argument of the adverbial binds the situation argument of the nominal subject, leading to the interpretation indicated in (28c). Since the subject is interpreted with respect to a new (but anchored) reference situation, definite DPs can only have a weak interpretation. This bleeds the necessity of entering into a relation with Fin^0 for the assignment of a context value for the individual argument of a subject. Thus, I conclude that subjects in the presence of frame adverbials do not need to be anchored by Fin^0 since they receive a bound interpretation.

28. a. Swans are white.
    b. In Australia, swans are black.
    c. G_x in Australia (s_1) and swans (x, s_1) $\rightarrow$ black (x)

In the following section, we will see that the pattern of subject licensing in Cimbrian is replicated in systems of complementizer agreements in West Germanic dialects.

7. Complementizer Agreements in West Germanic

As van Koppen (2016) showed, complementizer agreement (CA) is a complex and manifold issue in West Germanic. This is illustrated in (29–31). If a subject is focused, some dialects show CA, such as Austrian Bavarian in (29), and some dialects lack CA, such as Hellendoorn Dutch (30), and there are dialects in which a sentence is ungrammatical with or without CA, such as Frisian (31).

29. Warum-st grod DU mein Freind net griasst ho-st, vasteh i a net.
    Why-2P.SG PRT you my friend not greeted have-2P.SG. understand I too not
    Why you of all people didn't greet my friend, I don't understand either.'
    (Bavarian, Gmunden dialect, Gruber 2008, p. 53)
30. dat/* darr-e [zölfs wiej] de wedstrijd wint
    that/that-Agr even we the game win (Hellendoorn Dutch, van Koppen 2012)
31. a. * Hy leaude dat-st moarn do komme soest
       he believes that-2P.SG tomorrow you come should-2P.SG
    b. * Hy leaude dat moarn do komme soest
       he believes that tomorrow you come should-2P.SG
       (Frisian, Germen de Haan p.c., Fuß 2008, p. 85)

van Koppen (2012) argued that there are two types of CA to be distinguished. In a type A dialect, such as the dialect Tegelen Dutch, an agreement suffix is similar to an agreement suffix on a verb and CA is insensitive to subject movement and to subject modification. In a type B dialect, such as Hellendoorn Dutch, an agreement suffix differs from an agreement suffix on a verb, displaying the phenomenon of so-called double agreement (*DA*). Furthermore, in these dialects, the agreement suffix is of pronominal origin and CA is sensitive to subject movement and subject modification. I will illustrate the variable nature of *DA* that occurs in various West Germanic dialects with van Koppen's (2012) data from Hellendoorn Dutch.

It is illustrated in (32) that if a subject is moved into a preverbal position, *DA* (-e) is ruled out, while if a subject stays in a lower position, as in a yes/no question that requires a V1-order, *DA* is necessary. It is illustrated in (33) that if a subject is modified by a focus particle, *DA* is ungrammatical, and (34) shows that in the presence of a frame adverbial, *DA* is excluded, while (35) shows that if a focused subject is moved to a higher position and has a definite reading, as is the case for the first person pronoun *wiej, DA* is again excluded.

32. a. Wiej binn-t /*binn-e den besten!
    We are the best!
    b. Binn-e /*binn-t wiej den besten?
    Are we the best?
33. dat/* darr-e [zölfs wiej] de wedstrijd wint
    that/that-Agr even we the game win
34. dat/* darr-e [op den wärmsten dag van't joar] wiej tegen oonze wil ewärkt hebt
    that/that-Agr on the warmest day of the year we against our will worked have
35. WIEJ denkt Jan dat/*darr-e die pries ewönnen hebt, niet ZIEJ
    we think Jan that/that-1P.SG that prize won have, not they
    'WE John thinks won that prize, not THEM.'

(Hellendoorn Dutch, van Koppen 2012, p. 138)

Let us now have a look at the present accounts of CA in West Germanic. Because of the special properties of *DA*, two types of accounts have been proposed in the literature. For dialects of type A, where neither movement of a subject nor subject modification have an effect on the appearance of CA, it is assumed that CA is based on an agree relation between a C-head and a subject. For type B dialects, Fuß (2016) proposes a prosodic account in terms of string adjacency. In particular, Fuß (2016) assumes the post-syntactic movement of agreement features, which depends on strict string adjacency between the subject and Fin_0, accounting for the intervention effect of frame adverbials, as in (33) above. However, such an account cannot be extended to the dialects of type A (in which CA occurs in the presence of intervening elements), as was argued by Haegeman and van Koppen (2012), and the agreement-based account cannot be extended to dialects of type B. Thus, neither account can explain the phenomena in all of the diverse languages/dialects that display CA.

Before we sketch an alternative account that explains CA in both types of dialects, let me provide a description of the relevant data for CA in relation to the facts of subject-licensing in Cimbrian. It is immediately clear that the data in Cimbrian and CA in West Germanic exhibit very similar patterns that arguably call for a unified account. Let us thus consider whether the facts of CA can be explained in terms of anchoring a subject to a context. This implies that CA agreement should be seen as an alternative anchor (of the subject) such as clitics and *da* in Cimbrian.

First, I note that if the position of the complementizer (and the Wackernagel position) is identified with Fin^0 in the C-domain, then an alternative anchor is expected since the subject is prevented, for syntactic reasons, from moving into [Spec, FinP]—since no element can precede a complementizer in Germanic—to anchor the clause in embedded clauses. For matrix clauses, as shown in (31) and (34), where a subject moves into or through [Spec, FinP], no alternative anchor is needed.

Furthermore, I note that CA in Hellendoorn Dutch appears to behave similarly to subject clitics in Cimbrian (rather than *da*) in being incompatible with focus. This is in line with the observation that an agreement morpheme in a type B dialect is related in form to subject pronouns. Finally, (34) illustrates that a frame adverbial in a type B dialect bleeds CA since frame adverbials, as I discussed above, allow for the anchoring of a subject without any clitic via binding.

In conclusion, I make the following proposal: by considering referential anchoring, a uniform syntactic account becomes feasible where the two types of CA are related by a diachronic process of reanalyzing movement as agreement (cf. Wratil 2016 on *DA* in Carinthian and Kansas Bukovina Bohemian) along the following lines:

(A) Type B dialects involve the movement of a subject clitic or of a correlate of the subject into Fin^0 to anchor the predicate. This is why in cases of subject movement and subject modification, no overt CA morpheme appears. Dialects may then be taken to differ as to whether they allow subject clitics with focused subjects, as I propose is the case in Austrian Bavarian (see the data in (29) above), or whether they resort to a silent adnominal *da*. Dialects may not allow subject clitics with focused subjects, but they may also lack an adnominal *da*, and then the sentence is expected to be ungrammatical with or without CA, as is the case in Frisian (see the data in (31) above).

(B) Type A dialects are characterized by the loss of *DA* that can be explained in the present account in the following terms: the clitic is reanalyzed as verbal inflection, but the agree relation between Fin^0 and the subject remains and is interpreted as feature evaluation, as specified in (36) with the consequence that no intervention or modification effect is expected and the agreement morpheme is assimilated (or identical) to the agreement morpheme on the finite verb.

36. If term A agrees with function f (x), where x is a free variable, then x is evaluated with respect to the value assigned to A.

8. Conclusions

I have presented arguments showing that complementizer agreement is more than a quirky formal effect that appears in some West Germanic dialects, where we may ask what it is good for, and I have argued that CA can be taken to serve a purpose, namely, to anchor the subject in the context.

Furthermore, I have argued that the distribution of indefinite DPs and of strong and weak definite DPs in clauses follows from the assumption that these DPs must enter into a licensing relation with Fin^0 in the C-domain. These arguments are based on the assumption that the context values of discourse referents and the values pertaining to the utterance and reference situations become accessible in this position. This approach thus throws new light onto phenomena such as clitic movement to a high position in clauses that occur in many languages, as well as onto the scrambling of definite DPs into higher positions in the middle field (that may be identified with positions in the lower C-domain) in German. These operations then cease to be quirky formal properties of these languages, and they can be taken, as with CA, as we argued in the previous section, to serve to license definite DPs in the context.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable since the study did not involve human subjects.

Informed Consent Statement: Not applicable for studies not involving human subjects.

Data Availability Statement: Data are based on the author's own judgements or taken from publications cited and properly referenced.

Acknowledgments: I thank three anonymous reviewers for their comments on an earlier version of this paper.

Conflicts of Interest: The author declares no conflict of interest.

Notes

1 I dedicate this paper to the pianist Silvia Pezzotta and to the beauty of language and music that makes up the essence of our human nature. Interested readers are referred to https://www.youtube.com/watch?v=fCOKUUFOOyU for a taste of Silvia's artistry.

2 The bottleneck effect was introduced by Haegeman (1996) and Roberts (2004) to account for the V2 property in an extended C-domain, with the bottleneck assuring that, maximally, one constituent can be moved from the T-domain into the C-domain that hosts the finite verb in languages observing the V2 rule.

3 An anonymous reviewer pointed out that (18a) and (18b) constitute the following two quite different constructions: a low, not-raised subject which is obligatory and a right-dislocated subject where the subject DP is optional. I propose that both constructions derive from a low and extended subject containing a correlate: one is a focussed, stressed subject, and the other is a discourse anaphoric destressed subject. The right-dislocation in (18b) is necessary when the language, as occurs in Italian, does not allow for destressing in situ or for scrambling, as occurs in German.

References

Bidese, Ermenegildo, and Alessandra Tomaselli. 2005. Formen der 'Herausstellung' und Verlust der V2-Restriktion in der Geschichte der zimbrischen Sprache. In *Das Zimbrische zwischen Germanisch und Romanisch*. Edited by Ermenegildo Bidese, James R. Dow and Thomas Stolz. Bochum: Brockmeyer, pp. 71–92.

Bidese, Ermenegildo, and Alessandra Tomaselli. 2018. Developing pro-drop. The case of Cimbrian. In *Null Subjects in Generative Grammar. A Synchronic and Diachronic Perspective*. Edited by Federica Cognola and Jan Casalicchio. Oxford: OUP.
Bidese, Ermenegildo, Andrea Padovan, and Alessandra Tomaselli. 2012. A binary system of complementizers in Cimbrian relative clauses. In *Working Papers in Scandinavian Syntax*. Lund: Lund University, vol. 90, pp. 1–20.
Cognola, Federica, and Roland Hinterhölzl. 2020. Syntactic and semantic restrictions in the licensing of subjects in Cimbrian main clauses. *Linguistische Berichte* 263: 1–33. [CrossRef] [PubMed]
Diesing, Molly. 1992. *Indefinites*. Cambridge: MIT Press.
Donnellan, Keith. 1966. Reference and definite descriptions. *Philosophical Review* 75: 281–304. [CrossRef]
Ebert, Karen. 1971. Zwei Formen des bestimmten Artikels [Two forms of the definite article]. In *Probleme und Fortschritte der Transformationsgrammatik*. Edited by Dieter Wunderlich. Munich: Hueber, pp. 159–74.
Frege, Gottlob. 1892. Über Sinn und Bedeutung. *Zeitschrift für Philosophie und Philosophische Kritik* 100: 113–26.
Fuß, Eric. 2008. Multiple agreement and the representation of inflection in the c-domain. *Linguistische Berichte* 213: 77–106. [CrossRef]
Fuß, Eric. 2016. Complementizer agreement (in Bavarian): Feature inheritance or feature insertion? In *Bavarian Syntax. Contributions to the Theory of Syntax*. Edited by Günther Grewendorf and Helmut Weiß. Amsterdam: Benjamins.
Grewendorf, Günther, and Cecilia Poletto. 2015. Relative clauses in Cimbrian. In *Structure, Strategies and Beyond. Studies in Honor of Adriana Belletti*. Edited by Elisa Di Domenico. Amsterdam and Philadelphia: John Benjamins, pp. 393–416.
Gruber, Bettina. 2008. Complementizer Agreement: New Evidence from the Upper Austrian Variant of Gmunden. Master's thesis, University of Vienna, Vienna, Austria.
Haegeman, Liliane. 1996. Verb second, the split CP, and null subjects in early Dutch finite clauses. *Geneva Generative Papers* 4: 135–75.
Haegeman, Liliane, and Marjo van Koppen. 2012. Complementizer Agreement and the Relation between C0 and T0. *Linguistic Inquiry* 43: 441–54.
Heim, Irene. 1982. The Semantics of Definite and Indefinite Noun Phrases. Doctoral dissertation, MIT, Cambridge, MA, USA.
Heinrichs, Heinrich. 1954. *Studien zum bestimmten Artikel in den germanischen Sprachen [Studies on the Definite Article in Germanic Languages]*. Giessen: Wilhelm Schmitz.
Hinterhölzl, Roland. 2019. Subjects, Topics and Anchoring to the Context. *Syntax* 22: 2–3, Special Issue, 199–228. [CrossRef]
Hinterhölzl, Roland. 2024. Giving content to expletive *es* in German. *The Journal of Comparative Germanic Linguistics* 27: 1–31. [CrossRef]
Kolmer, Agnes. 2005. L'espletivo *da* come espletivo della posizione del soggetto enclitico pronominale nel Cimbro di Luserna. In *L'influsso dell'italiano sulla grammatica delle lingue minoritarie. Problemi di morfologia e sintassi*. Edited by Walter Breu. Rende: Centro Editoriale e Librario Università della Calabria, pp. 55–92.
Kratzer, Angelika. 1989. An investigation in the lumps of thought. *Linguistics and Philosophy* 12: 607–53. [CrossRef]
Milsark, Gary. 1974. Existential sentences in English. Doctoral dissertation, Massachusetts Institute of Technology, Cambridge, MA, USA.
Reichenbach, Hans. 1947. *Elements of Symbolic Logic*. New York: MacMillan Co.
Rizzi, Luigi. 1997. The Fine Structure of the Left Periphery. In *Elements of Grammar*. Edited by Liliane Haegeman. Kluwer International Handbooks of Linguistics. Dordrecht: Springer. [CrossRef]
Roberts, Ian G. 2004. The C-System in Brythonic Celtic Languages, V2, and the EPP. In *The Structure of CP and IP*. Edited by Luigi Rizzi. Oxford: Oxford University Press, pp. 297–328.
Russell, Bertrand. 1905. On denoting. *Mind* 14: 479–93. [CrossRef]
Scheutz, Heinrich. 1988. Determinantien und Definitheitsarten im Bairischen und Standarddeutschen [Determiners and definiteness types in Bavarian and Standard German]. In *Festschrift für Ingo Reiffenstein zum 60. Geburtstag*. Edited by Peter Stein, Andreas Weiss and Gerold Hayer. Göppingen: Kümmerle, pp. 231–53.
Schwager, Magdalena. 2007. (Non-)functional concepts: Definite articles in Bavarian. Paper presented at the Eighth Szkarska Poreba Workshop, Szrenica, Poland, February 23–27.
Schwarz, Florian. 2012. Situation pronouns in determiner phrases. *Natural Language Semantics* 20: 431–75. [CrossRef]
Stowell, Timothy. 1995. The phrase structure of Tense. In *Phrase Structure and the Lexicon*. Edited by Johan Rooryck and Laurie Zaenen. Dordrecht: Kluwer Publications, pp. 277–91.
van Koppen, Marjo. 2012. The distribution of phi-features in pronouns. *Natural Language and Linguistic Theory* 30: 135–77. [CrossRef]
van Koppen, Marjo. 2016. Complementizer Agreement. Master's thesis, University of Utrecht, Utrecht, The Netherlands.
Wratil, Melanie. 2016. The rise and fall of double agreement. In *Bavarian Syntax*. Edited by Günther Grewendorf and Helmut Weiß. Amsterdam: John Benjamins.
Zagona, Karen. 2013. Tense, aspect and modality. In *The Cambridge Handbook of Generative Syntax*. Part V, Syntax and the Internal Interfaces. Cambridge: CUP, pp. 746–92.

languages

MDPI

Article

Inferential Interrogatives with *qué* in Spanish

Ángel L. Jiménez-Fernández [1,*] and Mercedes Tubino-Blanco [2,*]

[1] Department of English Language, University of Seville, 41004 Sevilla, Spain
[2] Department of Spanish, Western Michigan University, Kalamazoo, MI 49008, USA
* Correspondence: ajimfer@us.es (Á.L.J.-F.); mercedes.tubino-blanco@wmich.edu (M.T.-B.)

Abstract: In this paper, we discuss the evidential properties of inferential interrogative sentences with *qué* in Spanish. This interrogative type exhibits the shape of a *wh*-question but the interpretation of a polar question. These sentences have the additional particularity that they are interrogatives with evidential material, which are attested but not frequent crosslinguistically, if compared with declarative evidentials. An interesting consequence of their double interrogative and evidential nature is the fact that both discourse participants have a prominent role in the interpretation of these sentences, as the Speaker makes the inference but the Addressee is requested for confirmation. To account for the construction, we assume a multiple-layered system that includes both Speech Act projection and Finiteness projection. In these two areas we simultaneously find evidential material housing the Speaker's inference, and a raised Addressee in its prominent interrogative position as the participant with the knowledge to provide the requested confirmation of the interrogative's truth value.

Keywords: evidentiality; evidential interrogatives; Speech Act Phrase; interrogative flip; confirmationals

Citation: Jiménez-Fernández, Ángel L., and Mercedes Tubino-Blanco. 2023. Inferential Interrogatives with *qué* in Spanish. *Languages* 8: 282. https://doi.org/10.3390/languages8040282

Academic Editor: Ana Ojea

Received: 16 October 2023
Revised: 20 November 2023
Accepted: 21 November 2023
Published: 30 November 2023

1. Introduction

The goal of this paper is to account for inferential interrogatives with *qué* in Spanish, an interrogative type with the appearance of a *wh*-question, since it involves the wh-pronoun *qué* 'what', but interpreted as an inferential yes/no question. Examples are shown in (1).[1]

(1) a. ¿Qué vienes, de la calle?[2]
what come:2SG from the street
'Are you coming back home (I infer)?
lit. what are you coming, back from the street?'

b. ¿Qué vas, en coche?
what go:2SG by car
'Are you going by car (I infer)?,
lit. what are you going, by car?'

This class of interrogatives has been previously identified as split interrogatives or split questions (Arregi 2007, 2010; Contreras and Roca 2007; López-Cortina 2003, 2009; Fernández-Soriano 2021), dislocated questions (Lorenzo 1994), and *wh*-doubling (Camacho 2002), since they seemingly involve a split in their structure. They have also been called compound interrogatives (Py 1971) and adjunct tags (Uriagereka 1988).

The reason why these interrogatives are identified as split is their hybrid nature. While their initial part exhibits the *wh*-pronoun *qué* 'what' as well as falling intonation, characteristic of *wh*-interrogatives, their final rising intonation and their interpretation as yes/no questions distinguish them from *wh*-interrogatives. A similar interrogative pattern has been found in Catalan (Contreras and Roca 2007) and English (López-Cortina 2009):

(2) a. What are you, on a diet?
b. What did you have, a food fight here? (López-Cortina 2009, p. 220 [1])

More recently, these constructions have been identified as invariable *qué* questions (Fernández-Sánchez and García-Pardo 2020a, 2020b, 2023; Reig Alamillo 2019) and non-matching split interrogatives (Fernández-Soriano 2021) to highlight the invariable nature of the interrogative pronoun involved in these constructions. Thus, although the tag contains a [wh] feature, it is crucially not a content *wh*-pronoun, hence, it is always realized as the default *wh*-word *qué* at PF, unlike other split interrogative classes. This explains the asymmetries between inferential interrogatives involving the default *wh*-operator *qué* and other split interrogative classes involving full-fledged *wh*-operators, such as *cómo* 'how', *cuándo* 'when', and *dónde* 'where' (3):

(3) a. ¿Qué llegaste, anoche?
what arrive:2SG.PERF last.night
'Did you arrive last night (I infer)?'

b. ¿Cuándo llegaste, anoche?
when arrive:2SG.PERF last.night
'When did you arrive, last night?'

Interpretation-wise, split interrogatives in general appear to have a confirmational value, i.e., the speaker requests information to confirm a previous suspicion or intuition (López-Cortina 2009). In this sense, these interrogatives show a strong evidential component, as also shown in Fernández-Sánchez and García-Pardo (2020b, 2023), since unlike other types of confirmation interrogatives such as tag questions, the expected reply in inferential interrogatives with *qué* is invariably constrained by the speaker's inferred or presupposed answer. For example, in the reading of a sentence such as (4), the speaker makes an inference about the truth value of the proposition on the basis of indirect evidence over the content of the proposition (e.g., the speaker sees the addressee while she enters the room with shopping bags).

(4) ¿Qué has ido, al supermercado?
what have:2SG gone, to.the supermarket
'Are you coming from the supermarket (I infer)?,
Lit. what have you gone, to the supermarket?'

The construction then shows an interpretative behavior similar to inferential evidentials, just as the one described by Bhadra (2017, 2018, 2020) for Bangla.

The evidential element in these constructions has an unusual Speaker-oriented interpretation, since interrogatives are typically Addressee-oriented. We propose that this unexpected interpretation follows from the interaction between the discursive elements present in the Finiteness Phrase (FinP) and the Speech Act Phrase (SAP) projections. More specifically, we propose a structure that involves the interplay between the presence of the Interrogative Flip, typical of evidential interrogatives (Aikhenvald 2004; San Roque et al. 2017), the evidential component present in these sentences and realized by the presence of the Speaker participant in a Fin projection (Bhadra 2020), and the presence, above ForceP, of an SAP where the Speaker and Addressee participants are anchored to the discourse and activate the inferential and confirmational interpretations, respectively, by means of a coindexation system with the relevant clausal elements (Bianchi 2003, 2006). This configuration explains why these interrogatives are interpreted as both inferential and confirmational (i.e., the speaker infers an answer in the tag, while the addressee is asked for full confirmation of the truth value of that inferred answer). This double discursive layer also accounts for the complex prosodic pattern typical of these constructions, in line with Escandell-Vidal's (2017) proposal for interrogatives with marked prosody.

The organization of this article is as follows. In Section 2, we list the grammatical properties displayed by inferential interrogatives with *qué* that were described in previous work such as Fernández-Sánchez and García-Pardo (2020a). In Section 3, we take a look at recent analyses of the construction. In Section 4, we discuss evidentiality in the context of interrogatives. Section 5 offers our proposal, a formal analysis based on the interaction between a Speech Act Phrase and a Finiteness Phrase. Section 6 addresses some consequences of our proposal and links our formal analysis with the grammatical properties of inferential interrogatives. Finally, in Section 7, we present the main conclusions.

2. Defining Characteristics of Inferential Interrogatives with *qué*

In this section, we review the main grammatical properties of inferential interrogatives with *qué* in Spanish, as they are mentioned in previous work (e.g., Fernández-Sánchez and García-Pardo 2020a), which contribute to establish crucial distinctions between these constructions and other types of interrogative clauses.

2.1. An Unexpected Intonation

Inferential interrogatives with *qué* exhibit a *wh*-pronoun in their initial part and falling intonation, characteristic of *wh*-interrogatives. What distinguishes these interrogatives from conventional *wh*-interrogatives is their final rising intonation and their interpretation as yes/no questions. This final part is frequently identified as the tag. A sentence such as (1a) above shows the intonation informally represented in (5):

(5) ¿Qué vienes, de la calle?

Compare with the rising intonation for the yes/no question in (6) and the falling intonation for the wh-question in (7); see Hualde (2005):

(6) ¿Vienes de la calle?

(7) ¿De dónde vienes?

A similar, but not identical, interrogative pattern has been found in Catalan (Contreras and Roca 2007) and English (López-Cortina 2009), as we indicated earlier:

(8) a. Què anirem, al teatre?
what go:2PL.FUT to.the theater
'What are you going, to the theater?'

b. Què ho faràs, al forn?
what CL do:2SG.FUT to.the oven
'What will you do it, in the oven?' (Contreras and Roca 2007, p. 145 [1])

(9) a. What are you, crazy?
b. What is he, your lawyer?
c. What are you, looking for a raise? (López-Cortina 2009, p. 220 [1])

What the sentences in (8–9) share with inferential interrogatives with *qué* is a complex prosodic pattern, as proposed in Escandell-Vidal (2017) also for other types of marked interrogatives.

2.2. The wh-Word Is Always qué 'What'

Lorenzo (1994), Fernández-Sánchez and García-Pardo (2020a, 2020b), and Fernández-Soriano (2021) identify crucial distinctions between the split interrogative class and inferential interrogatives with *qué*. A main distinction, which we also assume, is the fact that inferential interrogatives with *qué* exclusively involve the *wh*-word *qué* as in (10a), whereas

other split interrogatives involve any content *wh*-phrase, as illustrated in (10b). This is why Lorenzo (1994) treats *qué* in inferential interrogatives as an expletive *wh*-operator:

(10) a. Qué saludaste, a Pedro?
what greeted:2SG to oven
'Who did you greet, Pedro?, lit. what did you greet, Pedro?'

b. ¿A quién saludaste, a Pedro?
to who greeted:2SG to Pedro
'Who did you greet, Pedro?' (Fernández-Sánchez and García-Pardo 2020a, [19a])

2.3. The Operator qué Cannot Be Preceded by a Preposition

As pointed out by Fernández-Sánchez and García-Pardo (2020a), if the tag is a PP, the *wh*-word *qué* cannot be preceded by a preposition, in contrast with split questions with a content *wh*-pronoun (e.g., *dónde* 'where'), where a doubling preposition is obligatory. The contrast is shown in (11):

(11) a. ¿(*De) qué es, de Jaén?
from what is, from Jaén
'Is from Jaén that she is?, lit. what is she, from Jaén?

b. ¿*(De) dónde es, de Jaén?
from where is, from Jaén
'Is she from Jaén?'

The compatibility of the interrogative element with a preposition is a good test to distinguish these two types of split interrogatives, which is particularly useful when the full-fledged interrogative is *qué* 'what'. It is also an indicator that the *wh*-word in these constructions is not a full-fledged interrogative pronoun.

2.4. The wh-Question Is Not Independent

As Fernández-Sánchez and García-Pardo (2020a, 2020b) discuss, inferential interrogatives with *qué* cannot involve two independent interrogatives. While the initial part of split interrogatives is independent (12), that of inferential interrogatives with *qué* cannot stand on its own, as seen in the ungrammaticality of (13).[3]

(12) Split questions

a. ¿A quién saludaste?
to who greeted:2SG
'Who did you greet?'

b. ¿De dónde vienes?
from where come:2SG
'Where do you come from?'

c. ¿Dónde vas?
where go:2SG
'Where are you going?'

(13) Inferential interrogatives

a. *¿Qué saludaste?
what greeted:2SG
intended: 'Who did you greet?'

b. *¿Qué vienes?
what come:2SG
intended: 'Where are you coming from?'

c. *¿Qué vas?
what go:2SG
intended: 'How are you going?

This contrast is an indication that inferential interrogatives with *qué* are monoclausal. We develop this point further below in Section 3.3.

2.5. Inferential Interrogatives with qué Accept Tags Other than DP

Also reported in Fernández-Sánchez and García-Pardo (2020a), only inferential interrogatives with *qué* may be equivalent to a true yes/no question, and they may accept tags beyond the DP level, as seen in (14), with a DP followed by a comitative PP. Other split questions are more restricted in this sense.

(14) a. ¿Qué ha ido, Ana contigo?
what have:3SG gone Ana with.you
'Did Ana go with you (I infer)?'

b. *¿Quién ha ido, Ana contigo?
who have:3SG gone Ana with.you
Intended: who went, Ana with you?

Cf., ¿ha ido Ana contigo?
have:3SG gone Ana with.you
'Has Ana gone with you?'

2.6. Inferential Interrogatives with qué Have a Confirmational Flavor

Split interrogatives in general have a confirmational value, in the sense that the speaker requests the addressee to confirm a previous suspicion or intuition in their expected answer (López-Cortina 2009). This confirmational flavor is present in other types of interrogatives. For example, an inference based on evidentiality is precisely what we find in constructions such as Bianchi and Cruschina's (2019) polar interrogatives with fronted focus in English, also found in Spanish:

(15) a. Soup are you making? (Carter and McCarthy 2006, p. 780)

b. ¿Gazpacho estás haciendo?
Gazpacho be:2SG making
'Gazpacho are you making?'

This type of polar question has a confirmational value, which can only be understood if produced in a context that can be used to trigger the evidential reading. For example, the sentences in (15) are interpreted as confirmational if uttered when we enter the kitchen and see somebody preparing the necessary ingredients for the specific dish.

Interestingly, inferential interrogatives with *qué* show a behavior that is similar to inferential evidentials crosslinguistically. Bhadra (2018) analyzes the evidential marker *naki* in Bangla as a case of indirect evidence (Rooryck 2001) that can occur in different types of sentences, including interrogatives:

(16) Sita baRi giy-ech-e naki?
Sita home go-PERF-3SG NAKI
'Sita has gone home. Has she?
(Bhadra 2018, p. 1 [1a])

Bhadra (2018) claims that one of the roles of this particle is to ask for confirmation of the positive answer expected after inferring the truth-value from indirect evidence. In (17), we find another clear example from Bangla where the evidential marker *naki* indicates that some indirect evidence proves that what is asserted is true.

(17) Context: Ram knows that Mina has been thinking about going to America for a while now but has not made up her mind yet. Today, he suddenly sees several of her suitcases, all packed, sitting out in the hall and asks her brother:

Mina amerika chol-e ja-cche naki?
Mina America go-IMPERF go-3SG.PRES.PROGR NAKI
'(Given what I inferred) Mina is going away to America (is it true)?'
(Bhadra 2018, p. 2[3])

What is interesting about this particle is that this interpretation is only available if it appears in an interrogative sentence. If it appears in a declarative sentence, the interpretation of the particle does not have confirmation value, as seen in (18), which has a strictly reportative value.

(18) Context: Ram heard a rumor about his neighbor that he is now reporting to his friend Sita:

Mina naki	amerika chol-e	ja-cche
Mina NAKI	America go-IMPERF	go-3SG.PRES.PROG

'Mina is going away to America (I hear)'

(Bhadra 2018, p. 2[2])

The behavior of this particle is evidence of the projection of evidentiality material in syntax, as well as its composition interpretation, as will be argued in this paper. Sections 5 and 6 below further develop the idea that evidentiality projects in syntax.

2.7. Both Types of Split Questions May Be Preceded by Topics

Finally, both types of split questions allow the *wh*-word to be preceded by topics. Both sentences in (19) allow the preceding topic *el helado* 'the ice-cream'. As just seen, the sentence in (19a) is the inferential interrogative with *qué*, disallowing a preceding preposition, in contrast with the split interrogative in (19b), which does allow it:

(19) a. Inferential interrogative with *qué*

¿El	helado(,)	(*de)	qué es, de chocolate?
the	ice.cream	of	what is, of chocolate

'Is the ice-cream chocolate ice-cream (I infer)?'

b. Split question

¿El	helado(,)	de qué	es, de chocolate?
the	ice.cream	of what	is of chocolate

'Is the ice-cream chocolate ice-cream (I infer)?'

In the next section, we show recent analyses of the construction, divided between monoclausal and biclausal approaches.

3. Previous Recent Analyses

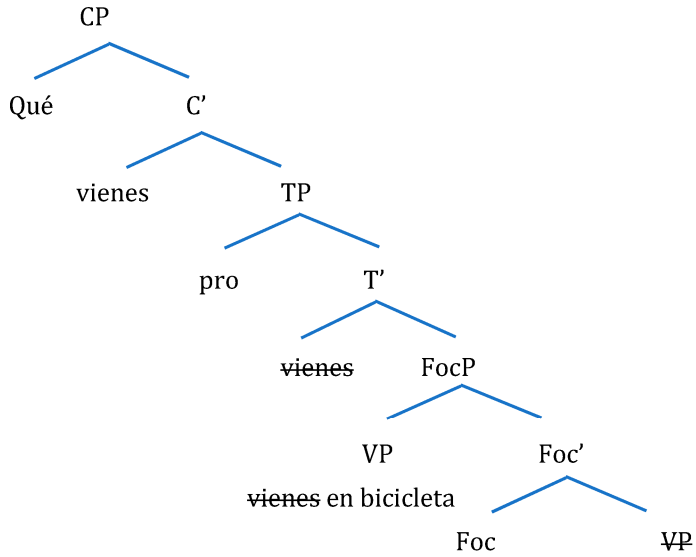
In this section, we discuss two recent analyses of Spanish inferential interrogatives with *qué*. One is monoclausal (Fernández-Sánchez and García-Pardo 2020a), whereas the other one is biclausal (Fernández-Soriano 2021).

3.1. A Monoclausal Analysis

Fernández-Sánchez and García-Pardo (2020a) suggest an analysis based on a low FocP (à la Belletti 2001, 2005), whereby the interrogative operator *qué* is base-generated in situ. More precisely, it originates in spec-CP, not involving any kind of movement. For the sentence in (20), the authors provide the analysis in (21):

(20) ¿Qué vienes, en bicicleta?
what come:2SG in bicycle
'Are you coming by bike (I infer)?' (Fernández-Sánchez and García-Pardo 2020a, 2020b)

(21)



Fernández-Sánchez and García-Pardo 's analysis is based on VP-movement to spec-FocP. However, this is problematic in cases in which we have a complex verb (with auxiliaries), since it is the auxiliary part that occupies T, whereas V remains lower. In such cases, we would obtain the ordering *Qué* Aux Subj VP: **¿Qué has tú venido, en bicicleta?*, contrary to facts. In particular, if the finite auxiliary is in T and then moves to C (as the lexical V *vienes* in (21)), the outcome involves the ordering Aux+Subject+VP, which is completely ill-formed.

Also, embedding of the inferential interrogative appears to be possible, at least in our dialect, which suggests that the operator undergoes long-distance movement to the matrix CP:[4]

(22) Long-distance movement

a. ¿Qué te fastidia que venga, el sábado?
what CL:2SG annoy:3SG that come:1SG, the Saturday
'Do you regret that I'm visiting on Saturday (I infer)?'

b. ¿Qué no te viene bien que venga, el sábado?
what NEG CL:2SG come:3SG well that come:1SG, the Saturday
'Isn't it good for you that I'm coming Saturday (I infer)?'

In addition, as shown by López-Cortina (2009), some Spanish varieties allow the interrogative pronoun *qué* to appear in situ in these constructions, and the same is found in English. Compare (23) and (24):

(23) a. Ecuadorian and Chilean Spanish
Vas qué, en tren?
go:2SG what on train?
'You go what, by train?'

b. Some varieties of American English
%You are going what, by train?

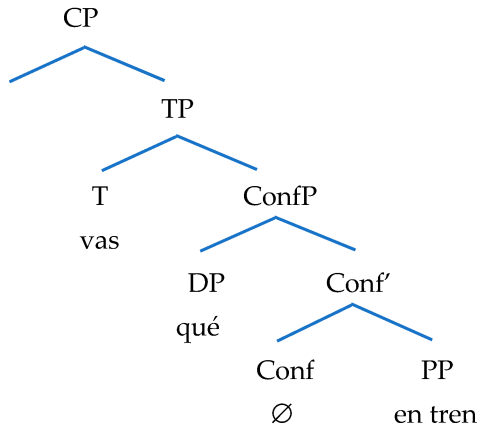
(24) a. ¿Qué vas, en tren?

b. What are you going, by train?

An analysis where the operator is base-generated is inconsistent with this data, as the alternation between in situ and *wh*-first constructions favor a movement analysis.[5] If the operator occupies [Spec, CP] from the beginning, the connection between (23) and (24) is lost. Note that their interpretation is identical, and the only difference is syntactic, i.e., no movement of *qué* in (23), and movement of this operator in (24).

The analysis proposed by López-Cortina (2009) is based on the projection of a Confirmation Phrase (ConfP), whose complement is the tag of the interrogative and whose specifier is the operator *qué*. In (25), we offer the analysis proposed by López-Cortina for sentences such as (24a):

(25)



Depending on the variety of Spanish, the operator will undergo *wh*-movement or remain in its original position.

This analysis captures the confirmational meaning of these interrogatives, but it fails to account for the focus interpretation of the tag. Actually, as is clear from the derivation in (25), no focus interpretation is taken into account by López-Cortina. In turn, Fernández-Sánchez and García-Pardo (2020a) propose that the tag is focused, as seen in their analysis in (21). In our analysis in Section 5, we also agree with a focus interpretation of the tag.

Below, in Section 3.3, we also argue in favor of a monoclausal analysis. In view of the data in this section, we also propose that the interrogative operator is the result of raising, as it appears to account for the behavior of these constructions regardless of the variety.

3.2. A Biclausal Analysis

Fernández-Soriano (2021), in turn, suggests the following biclausal analysis of her non-matching split interrogative clauses:

(26) [CP [Qué]$_i$ . . . [IP I . . . **[FP** t$_i$ **[F′ Ø** [CP tag$_j$ [~~IP. . . [I′ I [VP. . . t$_j$~~]]]]]]]]

In her analysis, the whole IP in the second clause is subject to ellipsis *à la* Merchant (2004). For Fernández-Soriano, the neuter operator *qué* and the tag are contained in an FP or Speech Phrase, which she claims is the phrase corresponding to discourse phrases in monoclausal analyses. In our proposal, we will see that the evidentiality reading of these interrogatives is the consequence of projecting a Speech Act Phrase, but this will be located in the top of the tree (Miyagawa 2017, 2022), not in the middle field.[6] In Fernández-Soriano's analysis, the evidential interpretation is the responsibility of FP in the highest CP in (26), whereas in our proposal this interpretation is the consequence of the projection of a Speech Act Phrase on top of the only CP.

A biclausal analysis is justified for split questions with content interrogative pronouns, since both parts are independently grammatical. This is, however, not what happens in the interrogatives under study here, whose first part is not grammatical on its own, as argued by Fernández-Sánchez and García-Pardo (2020a, 2020b). That is, these constructions do not consist of two independent interrogatives (i.e., a *wh*-interrogative followed by a yes/no question), since the *wh*-portion is not a well-formed independent clause in Spanish (27).

(27) *¿Qué vas?
what go:2SG
Intended: 'How are you traveling?, lit. what are you going?

cf. ¿Cómo vas?
how go:2SG
'How are you going?

The inferential interrogative in (27) cannot be used as an independent sentence, contrary to what a biclausal analysis predicts. Without a clear justification of the biclausal nature of these interrogatives, an ellipsis approach would be hard to maintain.

A monoclausal analysis where the verb moves to a focal position would be consistent with the data and the general behavior of verbs in Spanish interrogatives. We propose several tests next, which support a monoclausal analysis of inferential interrogatives with *qué*, while they hint at further asymmetries with other split interrogatives.

3.3. Testing the Biclausal/Monoclausal Analyses

Constituent preposing is a diagnostic associated with biclausality in Spanish (28) (examples from Sainz-Maza Lecanda and Horn 2015), whereas clitic climbing is associated with monoclausality (29).

(28) a. Biclausal
Mirando las olas, andaba por la orilla del mar
Looking at.the waves, walked by the shore of.the sea
'Looking at the waves, I walked by the seashore'

b. Monoclausal
*Estudiando para los exámenes, María anda cuando puede
Studying for the exams Maria walks when can:3SG
Intended: 'Maria is studying for her exams whenever she can'

(29) Biclausal
*Se viven peleando
CL:3.RECIPR live:3PL fighting
Intended: they live fighting
Cf. Viven peleándose

Inferential interrogatives with *qué* show monoclausal behavior, as they disallow constituent preposing, e.g., preposing of the tag, even with the non-elided constituent (30), and they do allow clitic climbing (31):

(30) *¿(Vienes) corriendo, qué vienes?
come:2SG running, what come:2SG
Intended: running is how you're coming?
Cf. ¿Qué vienes, corriendo?

(31) ¿Qué se lo quiere, comer en la cama?
what CL:REFL CL:3SG.MAS.ACC want:3SG eat in the bed
'What does he want, to eat it in the bed?'
Cf. ¿Qué quiere, comérselo en la cama?

In (32), sentences exhibiting a complex tag are disallowed regardless of the position of the clitic, whereas the simplex tag renders the sentence grammatical:

(32) a. *¿Dónde se lo quiere, comer en la cama?
where CL:RFLX CL:3SG.MAS.ACC want:3SG eat in the bed
'Where does he want to eat it in the bed?'

b. *¿Dónde quiere, comérselo en la cama?

c. ¿Dónde se lo quiere comer, en la cama?

These tests support a monoclausal analysis of inferential interrogatives with *qué*, and they also hint at further asymmetries with other split interrogatives. We conclude that a monoclausal analysis more accurately reflects the behavior of inferential interrogatives with *qué* than a biclausal analysis. In Section 6, we offer further arguments based on the prosodic contour of these constructions that a monoclausal analysis more accurately reflects the behavior of inferential interrogatives with *qué* than a biclausal analysis.

4. Evidentiality in Inferential Interrogatives with *qué*

Evidentiality has been extensively studied in languages with morphological evidential particles (Aikhenvald and Dixon 2003; Aikhenvald 2004). In these languages, evidential particles explicitly mark the source of information in a number of ways. According to Aikhenvald (2004), these particles vary in different languages, and they may encode that the information was reported by someone else, that the information was experienced first-hand by the speaker, sometimes visually, sometimes non-visually (e.g., through hearing or smelling), or by means of inference. Typically, evidential markers are obligatory and morphologically contrasted, depending on the type of source they specifically encode, as seen in (33) for Tariana, an Arawakan language spoken in Brazilian Amazon:

(33) Tariana (Arawakan)

a. Visual evidential (recent past) *-ka*
Juse ɨɾida di-manika-ka
José football 3SG.NF-play-REC.P.VIS
'José has played football (we saw it)'

b. Non-visual evidential (recent past) *-mahka*
Juse ɨɾida di-manika-mahka
José football 3SG.NF-play-REC.P.NONVIS
'José has played football (we heard it)'

c. Inferred evidential (recent past) *-nihka*
Juse ɨɾida di-manika-nihka
José football 3SG.NF-play-REC.P.INFER
'José has played football (we infer it from visual evidence)'

d. Assumed evidential (recent past) *-sika*
Juse ɨɾida di-manika-sika
José football 3SG.NF-play-REC.P.ASSUM
'José has played football (we assume this on the basis of what we already know)'

e. Reported evidential (recent past) *-pidaka*
Juse ɨɾida di-manika-pidaka
José football 3SG.NF-play-REC.P.REP
'José has played football (we were told)'

(Aikhenvald 2004, p. 3 [1.1, 1.2, 1.3, 1.4, 1.5])

Some authors have specifically explored the presence of evidential particles in interrogatives (Speas and Tenny 2003; San Roque et al. 2017; Bhadra 2017, 2018, 2020). According to San Roque et al. (2017), evidentials both provide information about the utterance and associate that information with the speech act participants, thanks to their perspectivizing function. For these authors, interrogative utterances also marked for evidentiality combine two facets of the expression of epistemicity in language: while evidentials express the source of information one (e.g., a discourse participant) has for a given proposition, interrogatives involve the speech act of questioning, by means of which information is requested that is unknown to the speaker.

According to these authors, this double function seems paradoxical, as one typically asks about things one knows little about. In other words, it would be a paradox for the same discourse participant to both request information about something and indicate the source of their knowledge. In fact, evidentials are incompatible or restricted with interrogatives in a number of languages (Aikhenvald 2004).

In languages in which evidentials are permitted along with interrogatives, identical evidential markers may contribute contrasted information depending on whether they appear in declaratives or interrogatives (San Roque et al. 2017; Bhadra 2017, 2018, 2020), which suggests that the interpretation of evidential particles is determined compositionally.

In this section, we discuss previous work on evidentials in interrogatives, paying special attention to the contribution that evidentials make to the syntactic composition of interrogatives, particularly the left-most left periphery. We also discuss how the presence of evidential material impacts the interpretation and markedness of interrogatives as well as the participants' point of view.

4.1. Change of Perspective in Interrogatives and the Interrogative Flip in Inferential Interrogatives

A general characteristic exhibited by interrogatives with morphologically overt evidentials is the presence of the *Interrogative Flip* (Tenny and Speas 2004; San Roque et al. 2017), a phenomenon by means of which the same evidential particle takes the speaker perspective in a declarative sentence, while it takes the perspective of the addressee in an interrogative. In the example in (34) from Duna, a Papuan language, the evidential affix *yarua* is addressee-oriented in the interrogative in (34A), but the same affix is speaker-oriented in the answer (a declarative sentence) in (34B):

(34) Duna (Papuan)
A: ko roro-yarua=pe
2SG hot-SENS=INTER
'Are you hot (you feel)?
B: no roro-yarua
1SG hot-SENS
'I am hot (I feel)' (San Roque et al. 2017)

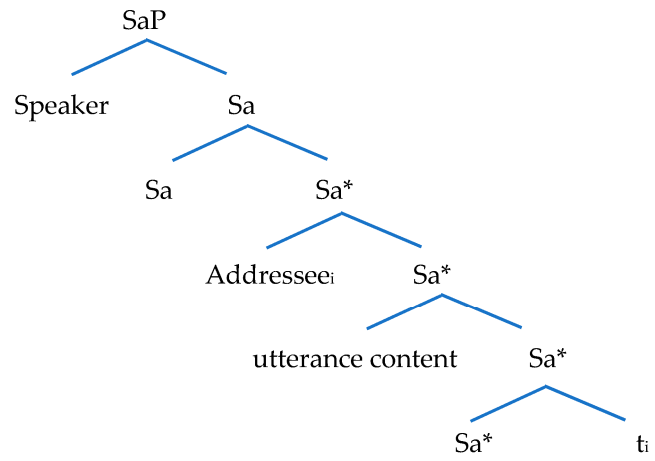
Besides Duna, this switch in perspective is obligatory in many world languages, including English (Speas and Tenny 2003; Tenny and Speas 2004), Japanese (Tenny 2006), Cuzco Quechua (Faller 2002), and Cheyenne (Murray 2010), to name a few, and, according to Bhadra (2020), it is associated with authority, in the sense that while the speaker has the authority in a declarative sentence in the sense that it is the speaker that possesses the knowledge behind an assertion, it is the addressee's knowledge that is sought in the answer to a question.

Speas and Tenny (2003) propose a system to explain syntactic structures attending to discourse participants (i.e., Speaker and Addressee) in the form of a set of syntactic projections housed in the left-most left periphery. This system is able to explain phenomena such as agreement with discourse participants instead of syntactic arguments, by means of coindexation, as seen in the case of unagreement in Spanish, shown in (35), in which the inflected verb *vamos* 'we go' shows first-person plural agreement with the Speaker along with the plural subject *los lingüistas* 'linguists'.

(35) Los lingüistas nos vamos de la sala
the.PL linguists CL:1PL go:1PL from the room
'We linguists are leaving the room'
(Jiménez-Fernández and Tubino-Blanco 2022, [33])

The diagram in (36) shows Tenny and Speas' (2004) proposed structure for an interrogative, which incorporates the Interrogative Flip:[7]

(36)



In the system proposed by Speas and Tenny (2003) and Tenny and Speas (2004), the *Seat of Knowledge*, which Tenny and Speas (2004) situate within the *Utterance Content*, is an evidential argument that stands for 'the sentient individual who is responsible for the truth of a proposition' (Tenny and Speas 2004). In the case of interrogatives, both the Utterance Content and Seat of Knowledge are controlled by the raised Addressee. This explains why the Addressee is the discourse participant that possesses the knowledge to provide the answer to the question. For example, the Seat of Knowledge is named by evidential verbs like *appear*. Because of the Interrogative flip, *appear* will be anchored to the Speaker in a declarative, but to the Addressee in an interrogative, as seen in (37):

(37) a. Martin appears$_S$ to have missed his exam. (The speaker knows)
b. Does Martin appear$_A$ to have missed his exam? (The addressee knows)

Inferential interrogatives with *qué* do not seem to pattern with prototypical interrogatives with evidentials in that they do not appear to present the Interrogative Flip, at least apparently, since the inference takes the Speaker's perspective rather than the Addressee's. For example, in a sentence such as (38), it is the Speaker that makes an inference about the Addressee's prior location:[8]

(38) ¿Qué vienes, de la piscina?
INTER come:2SG, from the swimming.pool
'Are you coming from the swimming pool (I infer)?,
lit. What are you coming, from the swimming pool?'

Inferential evidentials in Bangla (Bhadra 2017, 2018, 2020) also seem to lack the Interrogative Flip, as the Speaker perspective is maintained in the inference, as seen in (39):

(39) Mina amerika cho-e ja-cche naki?
Mina America go-IMPERF go-3P.PRES.PROG NAKI
'(Given what I inferred) Mina is going away to America (is it true)?'
(Bhadra 2018)

Both in Bangla and Spanish inferential interrogatives, the Addressee still has the information that will make the requested confirmation possible. This suggests that the Addressee does have a prominent role in the Speech Act projection, as a consequence of the Interrogative Flip. For this reason, the activation of the Interrogative Flip in inferential interrogatives with *qué* in Spanish is assumed in our proposal.

4.2. Evidential Projections

Four types of evidentials have been identified grammatically (Speas 2004). Different authors have proposed a compositional interpretation derived either from the different coindexations between participants and speech acts (Speas and Tenny 2003) or by means of four different projections (Cinque 1999; Speas 2004). Cinque (1999) proposes the hierarchy shown in (40) based both on the position evidential morphemes tend to exhibit within a word and adverb placement in a sentence.

(40) Cinque's (1999) hierarchy
Speech Act Mood > Evaluative Mood > Evidential Mood > Epistemological Mode

The Speech Act projection determines the type of Speech Act (e.g., interrogative), whereas the Evidential projection determines the source of the speaker's evidence of the truth of a proposition. Evaluative mood signals an evaluation made by the speaker and epistemological mode ascertains the degree in which a speaker is certain about a proposition (Cinque 1999; Speas 2004). In the case of morphemes, Cinque notes that speech act or speaker evaluation morphemes tend to appear further from the verb root than other morphemes, as shown in (41).

(41) Malagasi
matetika > efa > mbola > V (O) > tsara > tanteraka > foana > intsony > **ve**
generally already still well completely always anymore speech act
(Cinque 1999, p. 43 [207])

As for adverbs, he argues that adverbs such as *honestly* and *frankly* are associated with the Speech Act projection, adverbs such as *luckily* are associated with the Evaluative projection, and adverbs such as *obviously* are associated with the Evidential projection. This has an impact on the linear order that these adverbs present in a clause, as seen in (42–44), whereby evidential adverbs follow both speech act and evaluative adverbs, but precede epistemological adverbs.

(42) Speech act adverb *honestly* preceding evaluative adverb *unfortunately*
a. Honestly I am unfortunately unable to help you.
b. *Unfortunately I am honestly unable to help you.

(43) Evaluative adverb *fortunately* preceding evidential adverb *evidently*
a. Fortunately, he had evidently had his own opinion of the matter.
b. *Evidently he had fortunately had his own opinion of the matter.

(44) Evidential adverb *clearly* preceding epistemic adverb *probably*
a. Clearly John probably will quickly learn French perfectly.
b. *Probably John clearly will quickly learn French perfectly.
(Cinque 1999, p. 33)

Although Spanish word order restrictions generally differ from English, Spanish adverbs also appear to respond to Cinque's hierarchy in unmarked word order, at least regarding speech act adverbs and evidential adverbs as compared to other adverbs, as seen in (45–47).

(45) Speech act adverb *sinceramente* 'sincerely' preceding evaluative adverbial *por suerte* 'fortunately'
a. Sinceramente, por suerte no puedo ir.
sincerely fortunately NEG can:1SG go
'I sincerely can't fortunately make it'
b. ??Por suerte, sinceramente no puedo ir.

(46) Speech act adverb *sinceramente* 'sincerely' preceding evidential adverb *claramente* 'clearly'

a. Sinceramente, claramente ya no está entusiasmado.
sincerely clearly no.longer be:3SG enthusiastic:MASC
'Sincerely, he's clearly no longer enthusiastic.'

b. *Claramente, sinceramente ya no está entusiasmado.

(47) Evidential adverb *claramente* 'clearly' preceding epistemic adverb *probablemente* 'probably'

a. Claramente, este chico probablemente no aprobará el examen.
clearly this boy probably NEG pass:FUT.3SG the exam
'Clearly, this boy won't probably pass his exam.'

b. *Probablemente este chico claramente no aprobará el examen.

Different types of adverbs are oriented to different discourse participants in inferential interrogatives with *qué*. The Addressee, associated with the Speech Act, controls the Utterance Content and high adverbs such as *francamente* 'frankly,' *honestamente* 'honestly,' and *sinceramente* 'sincerely.' The Speaker, in turn, appears to be associated with the clausal level, controlling the reference of evidential adverbs such as *obviamente* 'obviously,' *claramente* 'clearly,' and *evidentemente* 'evidently.'

In the sentence in (48), the adverb *honestamente* 'honestly' obligatorily appears as a high adverb, and it is associated with the Addressee in the interrogative; that is, it is the Addressee's honesty that is being requested. In the case of the evidential adverb *evidentemente* 'evidently' in (49), it is lower and controlled by the Speaker; that is, it is the Speaker that makes the inference rather than the Addressee, even if the sentence is interrogative.

(48) Speech Act adverbs are addressee-oriented

¿Honestamente$_i$ qué vienes$_i$, de la fiesta?
Honestly what come:2SG from the party
'Honestly, are you coming from the party,
lit. honestly, what are you, coming from the party?'

(49) Evidential adverbials are speaker-oriented

¿Qué vienes$_i$ evidentemente$_{*i}$, del supermercado?
what come:2SG evidently from.the supermarket
'Are you evidently coming from the supermarket,
lit. what are you evidently coming, from the supermarket?'

The contrast in the anchoring of the different adverbs with different clause participants in (48–49) is evidence of the activation of the interrogative flip in Spanish evidential interrogatives with *qué*, demonstrating the Addressee's raising to a higher position where it controls high adverbs. It also proves the existence of a clausal layer controlled by the Speaker. We explain how the Speaker's point of view becomes active in these constructions next.

4.3. The Clause Logophoric Component: Point of View

Speas (2004) and Bianchi (2003, 2006) propose the existence of logophoric pronouns in a clause, representing the discourse participants' point of view. According to Bianchi (2003), every finite clause is anchored to the time of utterance or speech event (S), whereas non-finite clauses need to be anchored to the main clause. The speech time (S) is located in Fin, in the left periphery, as shown in (50).

(50) [Force [(topic*) [(Focus) [Fin [. . . Tense VP]]]

(Bianchi 2003)

Bianchi further proposes that only a finite construction encoding S may license person agreement, as it corresponds to a speech event, identified by Bianchi as the center of deixis, which includes the discourse participants as well as spatial and temporal coordinates determining finiteness. Sells (1987) distinguishes three distinct logophoric roles, which are relevant here, as they differentiate between different sources of information being reported: the source is the individual doing the reporting, the self is the individual whose mind is reported, and the physical point of view from which the report is made is the pivot.

Bianchi formally identifies the speech event as a Logophoric Center. Each Logophoric Center obligatorily projects an animate participant (i.e., the Speaker or Source), a typically optional Addressee in speech events, although the Addressee may be obligatory depending on the nature of the speech event (e.g., commands, questions). The Logophoric Center also contains temporal and spatial coordinates. This system is interesting because these logophoric pronouns can be coindexed, either with clausal arguments or with discursive participants, which affects phenomena such as agreement with discursive participants. For example, it would explain the agreement of the verbal features with the speaker together with a clausal argument in interrogative sentences such as (51), typical between a medical doctor and a patient (Jiménez-Fernández and Tubino-Blanco 2022).

(51) ¿Cómo estamos hoy?
how be:1PL today
'How are we today?'

The presence of the Logophoric Center controlling a Speech Act projection would also account for restrictions associated with the inclusive or exclusive interpretation of first-person plural pronouns and their interplay with information structure in Spanish. This fact was shown, for example, in Jiménez-Fernández and Tubino-Blanco (2022). Their system explains the restriction to use an overt first-person plural pronoun in an out-of-the-blue context, as seen in (52).

(52) A: What's the plan?
B: #Nosotros vamos a la playa (if intended as inclusive)
1PL go:1PL to the beach
'We're going to the beach'

(Jiménez-Fernández and Tubino-Blanco 2022, p. 157 [38])

Blain and Déchaine (2007) argue that operators may enter a clause at different levels (ie., vP, AspP, AgreeP, CP), leading to different semantic consequences. At the CP level, evidentials affect the Speech Act; at the vP level, they introduce the Speaker's perspective in the predicate. In Speas' (2004) system, a logophoric pronoun may not be the same as the prominent participant in the matrix Speech Act. For example, in (53) the Speaker would be Evaluator but someone different would be Witness and Perceptor:

(53) [pro_i SAP [pro_i EvalP [pro_j EvidP [pro_j EpisP]]]]

In Bangla (Mukherjee 2008; Bhadra 2017, 2018, 2020), the evidential morpheme *naki* changes its evidential contribution depending on the Speech Act in which it appears. If the morpheme appears in a declarative sentence and middle position, the morpheme has a reportative interpretation, but if it appears in an interrogative sentence and final position, it has an inferential interpretation. In this sense, it is similar to inferential interrogatives with *qué*. In fact, Mukherjee (2008) proposes that in the case of interrogatives, *naki* is a confirmation particle, an interpretation that has been previously proposed also for the tag portion in Spanish inferential interrogatives (López-Cortina 2003, 2009). Bhadra nonetheless proposes that, in both cases, *naki* is an indirect evidential and its interpretation is compositional. We see examples in (54):

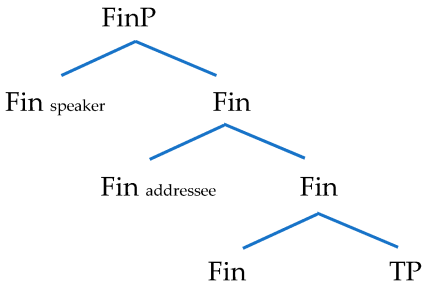
(54) Mukherjee (2008)

a. Shila naki gaan Sikh-ch-e
Shila INFER song learn-PROG-3
'Shila is learning music, as I have heard'

b. Sita baRi giy-ech-e naki?
Sita home go-PERF-3 CONFIR
'(From what I infer) Sita has gone home. Has she?'

Bhadra (2018) proposes two additional coordinates with respect to the ones proposed by Bianchi, which correspond with the discourse participants. These are different from the participants in Speas and Tenny's Speech Act projection. Bhadra proposes that either within or below the SaP projection we can find a FinP projection that accounts for the finite clause's point of view, which may correspond or not to the reference of discourse participants. This projection can be integrated either within SaP or below.

(55)



Keeping all these pieces in mind, we proceed to our analysis, which is spelled out in the next section.

5. A Formal Proposal for Inferential Interrogatives with *qué*

Our proposal is based on the following premises:

(i) The Addressee is located in a higher position within the Speech Act Phrase, which also explains the interpretation of the construction as an interrogative Speech Act (Miyagawa 2022). In this sense, it is the Addressee who controls the Utterance Content and Seat of Knowledge positions (Speas and Tenny 2003).

(ii) The Speaker maintains its coindexation with the evidential material by means of coindexation with FinP logophoric projection, where the confirmation proposition is located (Bhadra 2018).

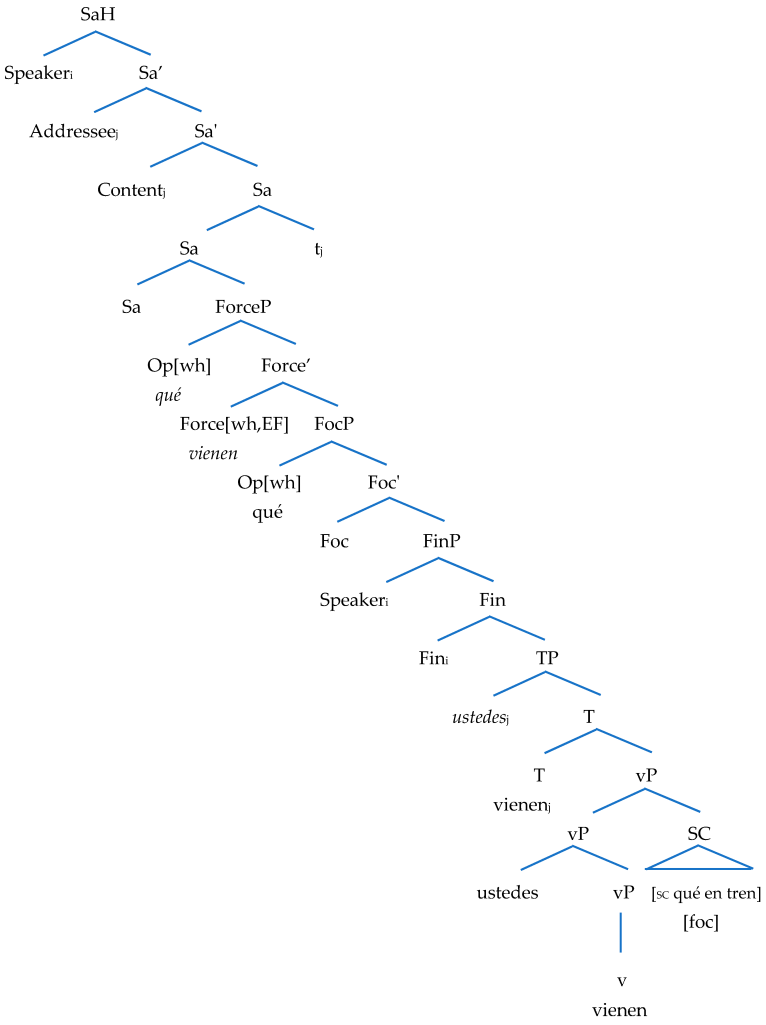
This explains why these sentences appear to be hybrid, in the sense that they appear to exhibit both partial and total interrogative behavior. It also explains why both Speaker and Addressee perspectives are present in the clause.

Evidence in favor of this hybrid structure is that the speech act adverb *honestamente* 'honestly' in (48) is anchored to the addressee, whereas the evidential adverb *evidentemente* 'evidently' is anchored to the speaker in (49). This leads us to propose two contrasted levels in the internal composition of these sentences, The Utterance level above is configured as an interrogative Speech Act, where the Addressee is the prominent role as a consequence of the Interrogative Flip, and where Speech Act adverbs are licensed. Below this level the clausal level can be found, where evidential adverbs are licensed and controlled by the Speaker, which represents this level's point of view (see Kim 2012 for a similar view and further argumentation on the anchoring of different types of adverbs by distinct discourse roles).

This configuration results in evidential interrogatives with *qué*, where we simultaneously find evidential material that houses the Speaker's point of view and classic interrogative material with the Addressee as the Seat of Knowledge, manifesting as the request for confirmation of the interrogative's truth value. To account for this configuration, we propose the construction in (56b) based on the sentence in (56a):

(56) a. ¿Qué vienen ustedes, en tren?
what come:2PL 2PL on train
'Do you all come by train (I infer)
lit. what do you all come, by train?'

b.



In the configuration in (56), the interrogative operator and the inferred material are base-generated in a Small Clause (SC) in *v*P. This accounts for the connection between the operator and the tag, given that the operator is clearly requesting information about the content of the tag. The inferential material remains in a lower position, which is typical of foci in Spanish. Both the evidential and discursive component of the structure take place at the CP layer, where the point of view and speech act relations take place. These evidential interrogatives are characterized by the fact that the evidential material is oriented to the Speaker rather than the Addressee. This is possible since it is the Speaker that occupies the [Spec, FinP] position, hence controlling the point of view of the material within the clause, including the tag (i.e., the SC). At the Speech act level, the interrogative character of the construction has additional consequences. The SC has a focus feature, which results in the movement of the operator *qué* to the [Spec, ForceP] position, passing through [Spec, FocP], since this ForceP projection has an Edge Feature (i.e., EPP in interrogatives). The reason why the SC has a focus feature is because one of its members (the tag) will end up being interpreted as the focus. The verb in turn moves to T and then to the ForceP head. At the discursive level, the Addressee takes a prominent place to reflect the interrogative character of the construction, moving from its neutral position in the complement of SaP to

a [Spec, SaP] position, controlling the Seat of Knowledge, typical of interrogatives, which will be interpreted in the form of an information-seeking utterance.[9]

This analysis reflects configurationally the structure of this marked interrogative type, in which the information sought by the Speaker is interpreted as a confirmation, as a consequence of the presence of evidential material controlled by the Speaker. Next, we discuss further structural and prosodic consequences resulting from this construction.

6. Some Consequences

In this section, we discuss both structural and prosodic consequences derived from the configuration proposed in (56).

6.1. Some Structural Consequences

In the proposed analysis in (56), we propose, contra Fernández-Sánchez and García-Pardo (2020a, 2020b), that both the operator and the verb move to Force. This would explain that when the subject is explicit, it is pronounced before the inference:

(57) ¿Qué vienen ustedes, en tren?
what come:2PL 2PL by train
'Do you all come by train (I infer)?
lit. what do you all come, by train?'

Also, embedding of the inferential interrogative appears to be possible, as seen in (22), here repeated as (58), which suggests that the operator undergoes long-distance movement to the matrix CP:

(58) Long-distance movement
a. ¿Qué te fastidia que venga, el sábado?
what CL:2SG annoy:3SG that come:1SG, the Saturday
'Do you regret that I'm visiting on Saturday (I infer)?'
b. ¿Qué no te viene bien que venga, el sábado?
what NEG CL:2SG come well that come:1SG, the Saturday
'Isn't it good for you that I'm coming Saturday (I infer)?'

In dialects without an Edge Feature in ForceP, both the operator *qué* and the inferential material would stay in situ, as in (59), which reflects the original position where the interrogative pronoun is generated:

(59) ¿Vas qué, en tren?
go:2SG what, by train
'Are you going by train (I infer)?
lit. are you going what, by train?'

(López-Cortina 2009, p. 221[5a])

Evidence that the tag is integrated within the clause rather than a second interrogative is the fact that the same temporal adverbial, associated with the matrix tense, may appear both with the tag or outside, as shown in (60). Note that the comma placement reflects the prosodic contour:

(60) a. ¿Qué vas mañana, al hospital?
what go:2SG tomorrow, to.the hospital

b. ¿Qué vas, al hospital mañana?
'Are you going to the hospital tomorrow (I infer)'

Further evidence in favor of the monoclausality of these sentences is the fact that the material in the tag can never be a finite clause, in contrast with other types of split interrogatives, as seen in (61):

(61) a. * ¿Qué va, va los fines de semana?
what go:3SG go:3SG the weekends
Intended: Does he go on weekends (I infer)?'

b. ¿Cuándo va, va los fines de semana?
when go:3SG go:3SG the weekends
'Does he go on weekends?'

Related to this point, Fernández-Sánchez and García-Pardo (2020b) argue that if these constructions were the result of ellipsis in the second clause, as a biclausal analysis of the construction would posit, sentences such as (62) would be grammatical, contrary to fact:

(62) *¿Qué estás, estás en Austin?
What be:2SG be:2SG in Austin
'Are you in Austin (I infer)?'

Fernández-Sánchez and García-Pardo (2020b)

Miyagawa and Hill (2023) argue that English interrogatives with evidential content may be compatible with *after all* despite the fact that this discourse marker is associated with declaratives, as shown in Sadock (1974):

(63) a. After all, your advisor is out of the country.
b. #After all, is your advisor out of the country?

(64) Evidential interrogative
After all, is the Pope catholic?

(Miyagawa and Hill 2023)

In spite of the interrogative nature of the inferential constructions studied here, they combine with the discourse marker *después de todo* 'after all', proof that these structures include evidential material at the syntactic level:

(65) Después de todo ¿qué vienes, con las manos vacías?
after of all what come:2SG with the:FEM.PL hand:PL empty:FEM.PL
'Do you come with empty hands after all (I infer)?'

Next, we discuss some prosodic consequences derived from the construction.

6.2. Prosodic Consequences

The analysis in (56) is in line with Escandell-Vidal's (2017) proposal according to which the evidential feature appears in interrogatives as a consequence of compositionality, resulting from the conjunction of point of view features and interrogative features. Escandell-Vidal distinguishes three contours associated with interrogatives: the canonical low-rise pattern (66) and two marked patterns, high-rise (67) and rise-fall (68).

(66) Low-rise yes/no question (canonical interrogative contour)
¿Has vivido siempre aquí?
have:2SG lived always here
'Have you always lived here?'

(67) High-rise (marked interrogative contour)
Cuando empezó la televisión en los años sesenta y tal pues me
when start:3SG.PERF the TV in the:PL years sixty and such well CL:1SG
parece muy bien que tenga que haber televisión, española

seem:3SG very well that have:3SG.SUBJ that there.be TV Spanish
pero **¿ahora?** // es que no le veo ningún sentido

but now be that NEG CL:3SG see:1SG no sense

'When television began in the sixties, it made sense to have a Spanish television,
but nowadays? I can't see the point of it!'

(68) Rise-fall (marked interrogative contour)

A:	¿Y si fueses	presidente de. . .?
	And if be:2SG.IMPERF.SUBJ	president of
	'If you were the president of. . .'	
B:	¿Si fuese	presidente de España?
	if be:1SG.IMPERF.SUBJ	president of Spain
	'If I were the president of Spain?'	

(Escandell-Vidal 2017)

The sentence in (66) shows a canonical yes/no interrogative whereby the Speaker asks some specific information that the Addressee is expected to provide. In (67), the Speaker asks a question the answer to which she already knows and is ready to provide. In (68), the interrogative is an echo-question. According to Escandell-Vidal, the canonical low-rise contour is the consequence of unspecified sentence polarity corresponding to the *wh*-operator, but the marked contours indicate the presence of evidential material in the sentence, indicating the information source. In her proposal, in the case of high-rise contours, the source of information would be the Self, that is, the evidential would be controlled by the Speaker. In the case of rise-fall contours, the information source would be the Other, that is, the evidential is addressee-oriented, according to Escandell-Vidal.

Although Escandell-Vidal does not discuss inferential interrogatives with *qué*, the marked fall-rise contour associated with these interrogatives is consistent with her proposal that marked intonation contours suggest the presence of evidential material. In this case, the source of information is hybrid, whereby the 'Self' (the Speaker) does the inferring, and the 'Other' (the Addressee) possesses the Seat of Knowledge.

For Escandell-Vidal (2017), the interrogative feature introduces a set of propositions {p, ~p}, where only one proposition is true, over which the evidential operates. The fall–rise contour associated with inferential interrogatives with *qué* indicates that the Seat of Knowledge is the 'Other', with scope over the true option and marked with the fall part of the contour. The inferential evidentiality part, associated with the 'Self' and rising intonation, operates over a hypothesis, which is explicit in the question.

Moreover, the intonational contour associated with these constructions is further evidence that they are monoclausal rather than biclausal constructions and that the tag is fully integrated within the interrogative, forming one prosodic unit, as also argued by Fernández-Sánchez and García-Pardo (2020b). Proof of this is that the tag cannot be an independent sentence, as shown also by Wiltschko and Heim (2016) for English confirmational tags:

(69) a. English confirmationals
You have a new dog. [*Eh?/*Huh?/*Right?]
Wiltschko and Heim (2016, p. 311[12])

b. Spanish inferential interrogatives with *qué*

*¿Qué	vienes?	¿De	la	calle?
what	come:2SG	from	the	street

'Are you coming back from outside (I infer)?
lit. ¿what are you coming? ¿from the street?'

7. Conclusions

Inferential interrogatives with *qué* are monoclausal interrogatives in which the interrogative operator moves to the CP area to validate [wh], [foc], and [EF] features. The interrogative operator is base-generated in a Small Clause within vP. The inference remains at the clausal level and takes the Speaker's perspective, thanks to the presence of a Speaker-oriented logophoric pronoun within FinP. This is not incompatible with the availability of the Interrogative Flip in these questions, which is addressee-oriented and occurs in the left-most left periphery. The consequence at the interpretive level is an unconventional interrogative clause with speaker-oriented evidential interpretation, the

appearance of a hybrid interrogative with both *wh*- and yes/no question behavior, and a marked fall–rise intonation.

Author Contributions: Conceptualization, Á.L.J.-F. and M.T.-B.; methodology, Á.L.J.-F. and M.T.-B.; software, Á.L.J.-F. and M.T.-B.; validation, Á.L.J.-F. and M.T.-B.; formal analysis, Á.L.J.-F. and M.T.-B.; investigatioÁ.L.J.-F. and M.T.-B.; resources, Á.L.J.-F. and M.T.-B.; data curation, Á.L.J.-F. and M.T.-B.; writing—original draft preparation, Á.L.J.-F. and M.T.-B.; writing—review and editing, Á.L.J.-F. and M.T.-B.; visualization, Á.L.J.-F. and M.T.-B.; supervision, Á.L.J.-F. and M.T.-B.; project administration, Á.L.J.-F. and M.T.-B.; funding acquisition, Á.L.J.-F. and M.T.-B. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Spain's Ministry of Science, Innovation, and Universities (MICINN), grant number PID2022-137233NB-I00.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are contained within the article.

Conflicts of Interest: The authors declare no conflict of interest.

Notes

[1] Glosskey: 1, 2, 3: 1st, 2nd, 3rd person; ASSUM: assumed; CL: clitic; COND: conditional; CONF: confirmational; FEM: feminine; FUT: future; IMPERF: imperfective; INFER: inferred; INTER: interrogative; MASC: masculine; NEG: negation; NF: non-final marker; NONVIS: non visual; PL: plural; PAST: past; PERF: perfective; PRES: present; PROG: progressive; REC.P: recent past; REP: reported; SG: singular; SENS: non-visual sensory; SUBJ: subjunctive; VIS: visual

[2] As a reviewer points out, the break or pause can also be made after the operator *qué*, as seen in Romanian:

(i) Ce, vii de pe stradă?
what come.2SG of on street
'lit. what, are you coming back from the street?'

In Spanish we also have this type of question. However, this is not the type of interrogative studied in this paper.

[3] In a similar vein, the tag in the inferential interrogative cannot be independent in an out-of-the-blue context. Hence, for the question (12b) i.e., *¿Qué saludaste, a Pedro?*, the tag *a Pedro?* cannot stand alone without a clear context.

[4] The sentences in (22) are the authors' judgments.

[5] Camacho (2002) and López-Cortina (2009) argue for a movement analysis of the operator *qué*. Camacho proposes a big DP where both the operator and the tag are generated; in a later step the operator undergoes movement to CP. Both Camacho's and López-Cortina's analyses are monoclausal.

[6] An ellipsis-based analysis is also proposed for Split Interrogatives by Arregi (2010). Recall that these interrogatives as similar to the inferential interrogatives studied here, but crucially the operator is a full *wh*-phrase in the former and the neuter *wh*-word *qué* in the latter:

(i) ¿Qué árbol plantó Juan, un roble?
what tree plant:3SG.PERF Juan an oak
'What tree did Juan plant, an oak?' (Arregi 2010, p. 540, example 1)

The ellipsis analysis is provided in (ii):

(ii) [CP what tree$_i$ planted Juan t$_i$] [CP an oak$_j$ ~~planted Juan~~ t$_j$]

In the second CP the verb and the subject are deleted, leaving only the preposed element which stands for the tag in the outcome.

[7] Speas and Tenny's original idea of having discourse categories such as SaP in the syntactic tree has been developed in Haegeman and Hill (2013); Miyagawa (2012, 2017, 2022); Portner et al. (2019); or Wiltschko (2014, 2017, 2021, 2022), among others.

[8] As a reviewer rightly points out, at least in some instances the addressee themselves provide certain types of evidence that facilitate the inference (e.g., the way they are dressed, wet hair, the addressee's routine, etc.). It is important to clarify that in these inferential interrogatives, the inference relies on contextual information that is inferred by the speaker. While contextual, this information is crucially not offered by the addressee, at least linguistically, and even if inferred by the speaker on the basis of the addressee's physical attributes, behavior, or habits known by the speaker, the inference still shows the speaker's perspective. It is important to distinguish between the inference made by the speaker (which is strictly the inference based on contextual information based on what the speaker observes about the addressee or any other kind of contextual information) and the actual information the addressee knows, which is what is requested by means of the confirmation component of these sentences.

[9] Because every utterance is delivered by a Speaker, the Speaker position appears in [Spec, SaP] by default.

References

Aikhenvald, Alexandra Y. 2004. *Evidentiality*. Oxford: Oxford University Press.

Aikhenvald, Alexandra Y., and R. M. W. Dixon. 2003. *Studies in Evidentiality*. Amsterdam: John Benjamins. [CrossRef]

Arregi, Karlos. 2007. Syntax and semantics of split questions. In *Romance Linguistics 2006: Selected Papers from the 36th Linguistics Symposium on Romance Languages (LSRL), New Brunswick, March–April 2006*. Edited by José Camacho, Nydia Flores-Ferrán, Liliana Sánchez, Viviane Déprez and María José Cabrera. Amsterdam and New York: John Benjamins, pp. 15–28. [CrossRef]

Arregi, Karlos. 2010. Ellipsis in split questions. *Natural Language and Linguistic Theory* 28: 539–92. [CrossRef]

Belletti, Adriana. 2001. Inversion as focalization. In *Subject Inversion in Romance and the Theory of Universal Grammar*. Edited by Afke Hulk and J.-Yves Pollock. New York: Oxford University Press, pp. 60–90.

Belletti, Adriana. 2005. Extended doubling and the VP periphery. *Probus* 17: 1–35. [CrossRef]

Bhadra, Diti. 2017. Evidentiality and Questions: Bangla at the Interfaces. Ph.D. thesis, Rutgers University, New Brunswick, NJ, USA.

Bhadra, Diti. 2018. Evidentials are syntax-sensitive: The view from Bangla. *Glossa: A Journal of General Linguistics* 3: 106. [CrossRef]

Bhadra, Diti. 2020. The semantics of evidentials in questions. *Journal of Semantics* 37: 367–423. [CrossRef]

Bianchi, Valentina. 2003. On finiteness as logophoric anchoring. In *Temps et Point de Vue/Tense and Point of View*. Edited by Jacqueline Guéron and Lilian Tamowski. Paris: Université de Paris X, pp. 213–46.

Bianchi, Valentina. 2006. On the syntax of personal arguments. *Lingua* 116: 2023–67. [CrossRef]

Bianchi, Valentina, and Silvio Cruschina. 2019. The derivation and interpretation of polar questions with a fronted focus. *Lingua* 170: 47–68. [CrossRef]

Blain, Eleanor M., and Rose-Marie Déchaine. 2007. Evidential types: Evidence from Cree dialects. *International Journal of American Linguistics* 73: 257–91. [CrossRef]

Camacho, José. 2002. Wh-doubling: Implications for the syntax of wh-movement. *Linguistic Inquiry* 33: 157–64. [CrossRef]

Carter, Ronald, and Michael McCarthy. 2006. *Cambridge Grammar of English*. Cambridge: Cambridge University Press.

Cinque, Guglielmo. 1999. *Adverbs and Functional Heads: A Crosslinguistic Perspective*. Oxford and New York: Oxford University Press.

Contreras, Joan M., and Francesc Roca. 2007. Un tipus especial d'oracions interrogatives: Les interrogatives escindides. *Caplletra* 42: 145–84.

Escandell-Vidal, Victoria. 2017. Intonation and evidentiality in Spanish polar interrogative. *Language and Speech* 60: 224–41. [CrossRef]

Faller, Martina T. 2002. Semantics and Pragmatics of Evidentials in Cuzco Quechua. Ph.D. thesis, Stanford University, Stanford, CA, USA.

Fernández-Sánchez, Javier, and Alfredo García-Pardo. 2020a. Invariable qué questions in Spanish. In *Language Patterns in Spanish and Beyond*. Edited by Juan J. Colomina-Almiñana and Sandro Sessarego. London: Routledge. [CrossRef]

Fernández-Sánchez, Javier, and Alfredo García-Pardo. 2020b. IQQs at the syntax-pragmatics interface. Paper presented at Linguistic Symposium on Romance Languages 50, University of Texas at Austin (online), July 7.

Fernández-Sánchez, Javier, and Alfredo García-Pardo. 2023. Teaching special questions: The role of semantics and pragmatics in colloquial interrogative structures in Spanish. *Hispania* 106: 45–66. [CrossRef]

Fernández-Soriano, Olga. 2021. The structure and interpretation of 'non-matching' split interrogatives in Spanish. In *Romance Languages and Linguistic Theory 2018: Selected Papers from 'Going Romance' 32, Utrecht*. Edited by Sergio Baauw, Frank Drijkoningen and Luisa Meroni. Amsterdam and Philadelphia: John Benjamins, pp. 97–116. [CrossRef]

Haegeman, Liliane, and Virginia Hill. 2013. The syntactization of discourse. In *Syntax and Its Limits*. Edited by Raffaella Folli, Christina Sevdali and Robert Truswell. Oxford: Oxford University Press, pp. 370–90.

Hualde, José Ignacio. 2005. *The Sounds of Spanish*. Cambridge: Cambridge University Press.

Jiménez-Fernández, Ángel L., and Mercedes Tubino-Blanco. 2022. On the interpretation of Spanish 1st person plural pronoun. In *Points of Convergence in Romance Linguistics*. Edited by Gabriela Alboiu and Ruth King. Amsterdam and Philadelphia: John Benjamins, pp. 143–159.

Kim, Sang Yoon. 2012. Una explicación formal de la orientación y distribución de los adverbios de actos de habla. *Revista Española de Lingüística* 42: 127–52.

López-Cortina, Jorge. 2003. The structure of split interrogatives. In *Theory, Practice, and Acquisition: Papers from the 6th Hispanic Linguistics Symposium and the 5th Conference on the Acquisition of Spanish and Portuguese*. Edited by Paula Kempchinsky and Carlos-Eduardo Piñeros. Sommerville: Cascadilla, pp. 140–55.

López-Cortina, Jorge. 2009. Split questions, extended projections, and dialect variation. In *Selected Proceedings of the 11th Hispanic Linguistics Symposium*. Edited by Joseph Collentine Maryellen García, Barbara Lafford and Francisco Marcos Marín. Somerville: Cascadilla, pp. 219–30.

Lorenzo, Guillermo. 1994. Qué expletivo en preguntas dislocadas. *Archivum* 44: 423–46.

Merchant, Jason. 2004. Fragments and ellipsis. *Linguistics and Philosophy* 27: 661–738. [CrossRef]

Miyagawa, Shigeru. 2012. Agreements that occur mainly in main clauses. In *Main Clause Phenomena. New Horizons*. Edited by Lobke Aelbrecht, Liliane Haegeman and Rachel Nye. Amsterdam: John Benjamins, pp. 79–112. [CrossRef]

Miyagawa, Shigeru. 2017. *Agreement Beyond Phi*. Linguistic Inquiry Monograph 75. Cambridge: MIT Press. [CrossRef]

Miyagawa, Shigeru. 2022. *Syntax in the Treetops*. Linguistic Inquiry 84. Cambridge: MIT Press. [CrossRef]

Miyagawa, Shigeru, and Virginia Hill. 2023. Commitment Phrase: Linking proposition to illocutionary force. *Linguistic Inquiry*, 1–15. [CrossRef]
Mukherjee, Aparna. 2008. Evidentiality: The Case of Bangla Naki. M. Phil dissertation, Jawaharlal Nehru University, New Delhi, India.
Murray, Sarah E. 2010. Evidentiality and the Structure of Speech Acts. Ph.D. thesis, Rutgers University, New Brunswick, NJ, USA.
Portner, Paul, Miok Pak, and Raffaella Zanuttini. 2019. The addressee at the syntax- semantics interface: Evidence from politeness and speech style. *Language* 95: 1–36. [CrossRef]
Py, Bernanrd. 1971. *La Interrogación en el Español Hablado de Madrid*. Bruxelles: Aimav.
Reig Alamillo, Asela. 2019. Las interrogativas hipotéticas con qué invariable en español: Un tipo de interrogativas parciales marcadas. *Studies in Hispanic and Lusophone Linguistics* 12: 155–77. [CrossRef]
Rooryck, Johan. 2001. Evidentiality, Part I. *Glot International* 5: 125–33.
Sadock, Jerrold M. 1974. *Toward a Linguistic Theory of Speech Acts*. New York: Academic Press.
Sainz-Maza Lecanda, Lorena, and Meagan E. Horn. 2015. The semantics and pragmatics of andar and venir + gerund. In *Hispanic Linguistics at the Crossroads: Theoretical Linguistics, Language Acquisition and Language Contact*. Edited by Rachel Klassen, Juana M. Liceras and Elena Valenzuela. Amsterdam and Philadelphia: John Benjamins, pp. 49–68. [CrossRef]
San Roque, Lila, Simeon Floyd, and Elisabeth Norcliffe. 2017. Evidentiality and interrogativity. *Lingua* 186: 120–43. [CrossRef]
Sells, Peter. 1987. Aspects of logophoricity. *Linguistic Inquiry* 18: 445–79.
Speas, Peggy. 2004. Evidentiality, logophoricity and the syntactic representation of pragmatic features. *Lingua* 114: 255–76. [CrossRef]
Speas, Peggy, and Carol Tenny. 2003. Configurational properties of point of view roles. *Asymmetry in Grammar* 1: 315–45. [CrossRef]
Tenny, Carol. 2006. Evidentiality, experiencers, and the syntax of sentience in Japanese. *Journal of East Asian Linguistics* 15: 245–88. [CrossRef]
Tenny, Carol, and Peggy Speas. 2004. The interaction of clausal syntax, discourse roles, and information structure in questions. Paper presented at Workshop on Syntax, Semantics and Pragmatics in Questions, ESSLLI 2004, Université Henri Poincaré, Nancy, France, August 13.
Uriagereka, Juan. 1988. On Government. Ph.D. dissertation, University of Conneticutt, Storss, CT, USA.
Wiltschko, Martina. 2014. *The Universal Structure of Categories: Towards a Formal Typology*. Cambridge: Cambridge University Press. [CrossRef]
Wiltschko, Martina. 2017. Ergative constellations in the structure of speech acts. In *The Oxford Handbook of Ergativity*. Edited by Jessica Coon, Diane Massam and Lisa Demena Travis. Oxford Handbooks Online. Oxford: Oxford University Press. [CrossRef]
Wiltschko, Martina. 2021. *The Grammar of Interactional Language*. Cambridge: Cambridge University Press. [CrossRef]
Wiltschko, Martina. 2022. Language is for thought and communication. *Glossa: A Journal of General Linguistics* 7: 1–19. [CrossRef]
Wiltschko, Martina, and Johannes Heim. 2016. The syntax of confirmationals: A neo-performative analysis. In *Outside the Clause*. Edited by Gunther Kaltenböck, Evelien Keizer and Arne Lohmann. Amsterdam: John Benjamins, pp. 305–40. [CrossRef]

languages

MDPI

Article

Locative Inversion in Old English Embedded Clauses

Sergio López-Martínez

Department of English, French and German Philology, University of Oviedo, 33011 Oviedo, Spain; lopezmsergio@uniovi.es

Abstract: A grammatical construction resembling Present-Day English locative inversion has already been found in Old English, with a fronted prepositional phrase prompting V2 word order, both in main and subordinate clauses. It has been demonstrated that several discourse-related factors influence the positioning of objects, fronted locatives, finite verbs and subjects in subordinate clauses. One of the main aims of the present paper is to provide a quantitative and qualitative analysis of the locative inversion construction in Old English subordinate clauses. The Old English data for this study were obtained from the *York–Toronto–Helsinki Parsed Corpus of Old English Prose*, and they were analysed using *Corpus Studio*. The results were compared with those for main clauses, and discourse-related factors such as PP anaphoricity or subject type were analysed in order to find the motivation for the existence of this alternation of word orders. PP anaphoricity proved not to be a determining factor in triggering finite verb inversion, while other factors such as subject weight and subject type do seem to motivate finite verb inversion, thus yielding an embedded PP-V-S word order.

Keywords: Old English; locative inversion; historical linguistics; syntax; information–structure

Citation: López-Martínez, Sergio. 2024. Locative Inversion in Old English Embedded Clauses. *Languages* 9: 171. https://doi.org/10.3390/languages9050171

Academic Editor: Ana Ojea

Received: 15 December 2023
Revised: 30 April 2024
Accepted: 1 May 2024
Published: 8 May 2024

1. Introduction

A great amount of the recent work on Old English word order focuses on information–structural factors as well as on syntactic ones (Van Kemenade and Los 2009; Biberauer and van Kemenade 2011; Van Kemenade and Milićev 2012; Taylor and Pintzuk 2012; Dreschler 2015). The construction known as locative inversion in present day languages has also received particular attention (cf. Bresnan 1994; Culicover and Levine 2001; Rizzi and Shlonsky 2006; Ojea 2020). In this type of construction, an adverb or prepositional phrase (PP) expressing location appears in the leftmost position of the clause, followed by the verb and with the subject postposed after the verb. It is assumed that locative inversion in Present-Day English (PDE) is only possible with intransitive, unaccusative verbs, i.e., verbs whose grammatical subject is not a semantic agent (cf. Burzio 1986; Hale and Keyser 2002), and it seems to work as an information rearranger (Dreschler 2015, p. 243) with a presentational function (Ojea 2020).

If we focus on Old English, Dreschler (2015) provides a diachronic study which includes a detailed syntactic and information–structural analysis of PP initial main clauses in Old English, concluding that among the functions of clause-initial PPs are those of local anchoring, contrast and frame-setting (2015: 265). Concerning the positioning of subjects, Dreschler (2015) argues that there is a tendency for unaccusative verbs to trigger inversions, as exemplified in (1):

(1) ***Æfter*** ***his*** ***deaðe*** *foran* *eft* *Cartainienses* *an* *Sicilie* *mid* *scipum*
after his death went again Carthaginians to Sicily with ships
'after his death went again the Carthaginians to Sicily with ships'
(coorosiu, Or_4:5.91.29.1854)
(From Dreschler 2015, p. 247, her 113)

In relation to the fronting of discourse-old constituents in Old English, López-Martínez (2019) explores the syntactic and information–structural implications of a series of sub-

ordinate constructions with two types of clause-initial constituents, namely determiner phrases functioning as objects, as in (2) below, and prepositional phrases, as in (3). In both cases, the finite verb appears in the second position in the subordinate clause instead of in final position. That study, which focuses mainly on clause-initial objects, takes into account the discourse status of both the clause-initial element and the PP-V-S word order. In constructions like (2), with a clause-initial object and the verb in second position, we seem to be dealing with embedded topicalization, while in constructions like (3), with a clause-initial PP, finite verb inversion and the subject in the final position, subject extraposition appears as the most plausible explanation. Furthermore, most of these constructions with a PP-V-S structure display an adjunct of space or time in the initial position, resembling those constructions with locative inversion in Present-Day English.

(2)

swa	***hit***	*healdað*	***Grecas***
swa			
as	it	keep	Greeks

'as the Greeks keep it'

(æLS[Basil]:142.546) (López-Martínez 2019, p. 64, his 29.a)

(3)

Forðæm	eac	wæs	*ðæt ðe*	***beforan***	***ðæm***	***temple***	*stod*
Because	also	was	that	before	the	temple	stood
æren	***ceac***	***onuppan***	***twelf***	***ærenum***	***oxum***		
brass	cauldron	upon	twelve	brass	oxen		

'Because it also was that a brass cauldron upon twelve brass oxen stood before the temple'

(CP:16.105.1.687) (López-Martínez 2019, p. 66, his 33)

One of the main objectives of this paper is to provide a more detailed analysis of initial PPs in subordinate clauses, paying particular attention to the informational–structural implications of verb inversion and late subjects. By presenting a qualitative and quantitative study of subordinate constructions with clause-initial PPs, both with and without finite verb inversion, the present work explores the possible motivations for the existence of these types of word order in Old English. Looking at the discourse status of clause-initial PPs, their anaphoricity will prove not to be a determining factor to trigger finite verb inversion in this type of construction. On the contrary, other factors such as subject weight and subject type seem to motivate finite verb inversion, thus yielding an embedded PP-V-S word order.

The structure of this paper is the following. Section 2 provides a series of theoretical considerations on Present-Day English, and Section 3 focuses of the syntax of Old English as a V2 language. Section 4 includes a description of the materials and methods, and Section 5 provides the results of the main query. Section 6 provides a detailed quantitative and qualitative analysis of the syntactic and informational–structural status of clause-initial PPs, the finite verbs and their subjects, paying special attention to late subjects. Finally, Section 7 presents some conclusions.

2. Locative Inversion in Present-Day English

Levin and Rappaport Hovav (1995, p. 218) present what they consider to be the main properties of the construction known as "locative inversion" in Present-Day English, which they illustrate with (4) below:

(4) In the distance appeared the towers and spires of a town which greatly resembled Oxford. (L. Bromfield, *The Farm*, 124)

(Levin and Rappaport Hovav 1995, p. 218, their 1)

Among those properties, they mention its noncanonical "PP V NP" word order, the presence of a locative or directional PP in "preverbal position", and the fact that the verb must be intransitive and more particularly unaccusative. Those unaccusative verbs, according to them, include mainly verbs of appearance, existence, directed motion and

manner of motion (p. 220). Similarly, Webelhuth (2011) presents the following list of properties for the locative inversion construction:

(5) a. Unusual word order: [PPLocAUX*V LOG-SUBJ]
b. The main verb must be intransitive.
c. The sentence must not be negated.
d. The logical subject must not be an anaphoric pronoun.
e. The relative familiarity constraint.
f. The "Displaced speech" effect

(Webelhuth 2011, p. 83)

We can see how Levin and Rappaport and Webelhuth agree on the intransitivity of locative inversion. Concerning the relative familiarity constraint in (5e), Webelhuth (2011) refers to Birner (1996, p. 90) and the discourse constraint which implies that "[t]he preposed element in an inversion must not be newer in the discourse than the postponed element." As regards the "Displaced speech" effect, Webelhuth (2011, p. 86), building on Drubig (1988) and Bolinger (1977), defines it as an effect displayed by locative inversion which implies that the speaker or writer has "privileged sensory access to the situation that is described." The data retrieved prove that these properties apply to embedded clauses in Old English as well, with subjects conveying new information and clauses with a presentational sense in most cases.

There are, however, some restrictions concerning the availability of locative inversion in embedded clauses. Consider the following examples from Sasaki (1998):

(6) a. That Bill rushed into the Oval Office is believed.
b. *That into the Oval Office rushed Bill is believed.
c. It is believed that Bill rushed into the Oval Office.
d. It is believed that into the Oval Office rushed Bill.

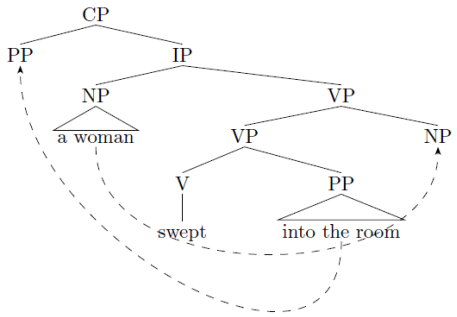
(Adapted from Sasaki 1998, p. 54, his 5)

Observing the evidence in (6), Sasaki proposes that locative inversion in embedded clauses is restricted to those which are in a CP selected by a bridge verb, as in (6c). Bridge verbs are those that allow for complementizer deletion (Van Kemenade 1997, p. 328). This restriction thus renders (6b) ungrammatical. Furthermore, Sasaki points out that locative inversion cannot take place in a clause without a complementiser, even after a bridge verb, as seen in (7) below. Old English data, however, prove that locative PPs can be fronted in embedded clauses in this language in contexts other than CPs selected by bridge verbs, including adverbial and relative clauses.

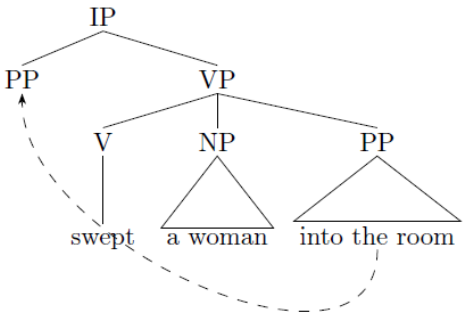
(7) a. Mary said that Bill rushed into the Oval Office.
b. *Mary said into the Oval Office rushed Bill.

Syntactically, authors such as Postal (1977, 2004), Chomsky (2008) and Ojea (2020) assume that the locative PP moves to CP and that there is a covert expletive in Spec-TP. Kim (2003, p. 3) presents the two main traditional analyses of locative inversion in terms of movement: the topicalization approach, as shown in (8) below (cf. Bowers 1976; Newmeyer 1987; Rochemont and Culicover 1990), and the unaccusative approach, illustrated in (9) (cf. Coopmans 1989; Hoekstra and Mulder 1990).

(8)



(9)



In the topicalization approach, the PP moves into a topic position as the specifier of CP, and the subject NP moves to a VP-adjoined position. On the other hand, in the unaccusative analysis, the NP is the object of an unaccusative verb and the locative PP moves into a subject position. If we assume the notion outlined in Fischer et al.'s work (2000, p. 109) that the non-final V position in the embedded clause bears no relation to topicalization, since it is a main-clause phenomenon, we could deem the topicalisation analysis problematic if we want to account for locative inversion constructions in Old English embedded clauses.

Reviewing the unaccusative approach, Kim (2003, pp. 7–9) indicates that the PP, which moves to the Spec-IP position, displays some subject properties. For instance, in relation to raising, we can find cases of a locative PP working as the subject of a raising verb, as is in (10) below.

(10) a. Over my windowsill seems to have crawled an entire army of ants.
b. On the hill appears to be located a cathedral.

(Kim 2003, p. 7, his 18)

Kim (2003, pp. 8–9) also argues that in the unaccusative approach, the postverbal NP displays object properties as it is generated as the object of an unaccusative verb. This is demonstrated, for example, by the fact that adverbs cannot appear between the verb and the NP as the NP is generated as the complement of the verb, and nothing can intervene between them, as seen in (11):

(11) a. In front of us walked Dana proudly.
b. *In front of us walked proudly Dana.

(Kim 2003, p. 9, his 28)

We must now consider the issue of the left periphery of the clause in relation to the syntax of Old English as a V2 language and clause-initial PPs.

3. V2 in Old English

From a syntactic point of view, Old English has traditionally been regarded as a V2 language, with the finite verb generally following an initial constituent in the main clauses (Fischer et al. 2000, p. 15). Fischer et al. distinguish different word order patterns for V2 sentences (2000: 105–108). The subject commonly appears as the first constituent of the main clause, with the finite verb following it, as seen in (12) and (13) below:

(12) **We habbað** hwæðere þa bysne on halgum bocum
we have nevertheless the examples in holy book
'We have, nevertheless, the examples in the holy book'
(*ÆCHom* I, 33.474.33)

(13) **Se Hælend wearð** þa gelomlice ætiwed his leornung-cnihtum
the Lord was then frequently shown his disciples
'The Lord then frequently appeared to his disciples'
ÆCHom I, 15.220.21)
(From Fischer et al. 2000)

When the first constituent in a main clause is not the subject, the finite verb often follows it, resulting in subject–verb inversion. This is the case with the so-called "operators", i.e., question elements as in (14) below, the negative *ne* as in (15), and the adverbials *þa* and *þone* as in (16). Inversion can take place with both nominal and pronominal subjects.

(14) **Hwi** wolde God swa lytles þinges him forwyrnan?
why would God so small thing him deny
'Why should God deny him such a small thing?'
(*ÆCHom* I, 1.14.2)

(15) **Ne** sceal he naht unalifedes don
not shall he nothing unlawful do
'He shall not do anything unlawful'
(*CP* 10.61.14)

(16) **Þa** wæs þæt folc þæs micclan welan ungemetlice brucende
then was the people the great prosperity excessively partaking
'Then the people were partaking excessively of the great prosperity.'
(*Or* 1.23.3)

Verb–subject order can occur when the first constituent is a non-subject if the subject is a full determiner phrase (DP), as in (17), but if the subject is a pronoun, as in (18), inversion is not possible in most cases:

(17) On twam þingum hæfde God þæs mannes sawle gedodod
in two things had God the man's soul endowed
'With two things God had endowed man's soul'
(*ÆCHom* I, 1.20.1)

(18) Forðon we sceolan mid ealle mod & mægene to Gode gecyrran
therefore we must with all mind and power to God turn
'Therefore we must turn to God with all our mind and power'
(*HomU*19 (BIHom 8) 26)

As regards embedded clauses, most authors such as Fischer et al. (2000, pp. 107–9) agree on the fact that the movement of the finite verb is more restricted and consider preposed constituents such as topics and question elements followed by the finite verb to be "a main-clause phenomenon." It is accepted, however, that the subject in embedded clauses always precedes the finite verb, "except in special constructions such as passives", and that "regular topics followed by the finite verb as in main clauses do not appear in

this position." Generally speaking, they illustrate the canonical S-Vfin-XP-V word order of embedded clauses, as shown in (19) below:

(19) Þæt **hi** mihton swa bealdlice Godes geleafan **bodian**
that they could so boldly God's faith preach
'that they could preach God's faith so boldly'
(ÆCHom I, 16.232.23)

In relation to V2 in embedded clauses, Salvesen and Walkden (2017) refer to the difference between CP-V2 and IP-V2 languages: in a CP-V2 language, the landing site for the finite verb is C^0 (via I^0), while in an IP-V2 language, it does not move any higher than I^0 (cf. Van Kemenade 1997; Pintzuk 1991; Kroch et al. 2001). Salvesen and Walkden (2017) propose what they label the "split hypothesis" based on Travis (1984, 1991) and Zwart (1991, 1993): it presupposes that the position of the finite verb depends on the nature of its preceding XP, i.e., the finite verb rises to C^0 when the first constituent of the clause is "a nonsubject", whereas it remains in Spec-IP when the subject is in the initial position (Salvesen and Walkden 2017, p. 170).

Salvesen and Walkden (2017, p. 173) argue that CP-V2 languages, or asymmetric V2 languages, can be divided into those that prohibit embedded V2 "whenever the complementizer is present", such as German, and those which allow embedded V2 "with an overt complementizer only in specific contexts", such as Mainland Scandinavian. Those contexts in which embedded V2 is allowed are usually complement clauses of the so-called 'bridge' verbs. In their study, Salvesen and Walkden (2017) find "only a handful of non-accidental counterexamples" of embedded V2 in Old English, which leads them to affirm that embedded V2 in Old English is completely ruled out.

Haeberli (2001, pp. 204–5), among others, assumes two potential landing sites for finite verbs (C with clause-initial operators like *þa* 'then' and Agr with clause-initial non-operators, like DPs or PPs), as well as two positions for subjects (a high position reserved for pronominal subjects and a lower position for full DP subjects), as illustrated in (20) below. To account for the asymmetry found in subordinate clauses, in which the canonical word order is V-final, Haeberli (2001, p. 223) assumes that the complementiser in this type of clause is generated in Fin and that the finite verb moves only to T, as summarized in (21):

(20) [CP [XP] C [AgrP **SU1** Agr [TP **SU2** . . .]]]
(21) [C [**SU1** [TP **SU2** (. . .) V (. . .)]]]
(Adapted from Haeberli 2001)

However, Biberauer and van Kemenade (2011, p. 18) suggest that the structure in (20) and (21) does not account for a series of informational–structural factors, which are particularly relevant for the placement of subjects. In order to reflect such factors, Biberauer and van Kemenade propose the structure illustrated in (22) below, with the higher subject position located within an articulated complementizer phrase (CP). We must note that higher subject positions are reserved for discourse-old subjects (and also pronominal objects), while the lower position is reserved for discourse-new subjects. This supports the idea in Pintzuk (1993) and López-Martínez (2019) that clause-initial elements other than the subject could occupy a higher position in the clause in Old English, even in subordinate clauses. If we are to assume this analysis applies to embedded clauses with fronted PPs, it is precisely that higher position which would be occupied by the fronted locative, thus providing a generative solution for this kind of construction.

(22) [CP XP C [FP **Discourse-Old Subj (SU1)** F [TP **Discourse-new Subj (SU2)** T . . .]]]
(Adapted from Biberauer and van Kemenade (2011)

Focusing on those clause-initial PPs in Old English main clauses, Dreschler (2015) shows that they generally function as local anchors, with additional functions such as

contrast and frame-setting. It seems that in main clauses, both the type of verb and the type of PP influence verb inversion in constructions which resemble locative inversion in PDE, and the information status of subjects is considered to be "a less important factor for subject placement" (Dreschler 2015, p. 242). Thus, inversion after clause-initial PPs occurs with unaccusative verbs, ascribed to the fact that unaccusative subjects originate as the internal argument in the VP and remain in that position. The present study tests whether this is the case for clause-initial PPs in embedded clauses with finite verb inversion.

4. Materials and Methods

As mentioned earlier, the present paper aims at providing a quantitative and qualitative study of subordinate clauses with clause-initial PPs in Old English, both with and without finite verb inversion (including main verbs, be, have and modals) and to explore the possible informational–structural factors which may motivate these types of word order. To do so, an Old English dataset was compiled from the *York–Toronto–Helsinki Parsed Corpus of Old English Prose* (Taylor et al. 2003), a corpus of 1.5 million words comprising 100 Old English prose texts of all types and genres. The search engine used to analyse the dataset was *Corpus Studio* (Komen 2009). For the sake of replicability, the code used for the query is included in the Appendix A.

Making use of the software Cesax (Komen 2012), the data obtained by *Corpus Studio* (Komen 2009) generated an Excel file with a list of every individual example, which discriminated each type of embedded word order and allowed for a fine analysis of each type of preposed PP, subject and verb type.

5. Results

The query submitted included two types of word order[1]: subordinate clauses with a clause-initial locative PP and the finite verb in final position (henceforth, PP-S-V word order), as seen in (23) below, and subordinate clauses with a clause-initial PP and the finite verb in second position (henceforth, PP-V-S), as repeated in (24). Temporal PPs were also included, since examples containing this type of element also fell under the category of presentational shared by locatives.

(23) Sua sua ***mid*** ***liðre*** ***wisðlunga*** ***mon*** *hors* *gestilleð*
So that with softer whistles man horses calm
'so that the horses are calmed still with softer whistles'
(CP:21.161.12.1098)

(24) Forðæm eac wæs *ðæt ðe* ***beforan*** ***ðæm*** ***temple*** *stod* ***æren***
Because also was that before the temple stood brass
ceac ***onuppan*** ***twelf*** ***ærenum*** ***oxum***
cauldron upon twelve brass oxen
'Because it also was that a brass cauldron upon twelve brass oxen stood before the temple'
(CP:16.105.1.687)

The query yielded a relatively even distribution of both types of word order, as illustrated in Table 1 below, with 255 tokens of embedded PP-S-V word order and 333 tokens of embedded PP-V-S word order. We must bear in mind that V2 is rare among all subordinate clauses with a PP in Old English, including those in which the PP follows the verb.

Table 1. OE query distribution.

PP-S-V	PP-V-S	Total
255 (43.4%)	333 (56.6%)	588 (100%)

Figure 1 below shows the evolution of both types of word order throughout the four sub-periods in which the *YCOE* divides the Old English period. We can observe that an

embedded PP-V-S word order (i.e. with finite verb inversion) was more numerous than that without inversion in the O2 and O3 sub-periods, particularly in O3, while the situation was inverted in the last sub-period, with embedded PP-S-V above its competing word order:

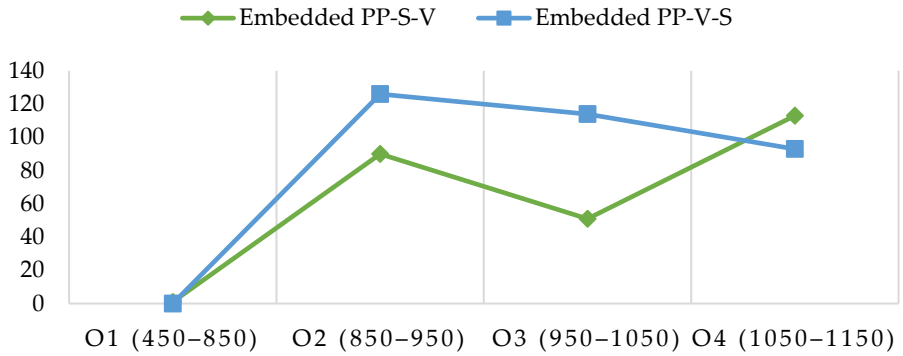


Figure 1. Initial PPs in embedded clauses through OE periods.

Bearing this in mind, the following section will examine the syntactic and informational–structural status of clause-initial locatives in these two types of configurations while also trying to assess the motivation for their distribution.

6. Discussion

As seen in the previous section, both embedded clauses with clause-initial locatives with and without inversion were abundant in Old English, with an apparently even distribution. Examples (23) and (24) above illustrate a clear difference in use between these two competing word orders: while (23) places the PP in the clause-initial position in the embedded clause for emphasis (i.e., it is with gentler whistles, *mid liðre wisðlunga,* that horses are calmed), with the subject 'man' *mon,* (24) is clearly presenting a scene, that of the front surroundings of a temple, following both the "displaced speech" effect described by Webelhuth (2011) for PDE and the relative familiarity constraint, with a subject conveying new information ('a brass cauldron upon twelve brass oxen').

This section now analyses the data obtained for the present study, paying particular attention to the status of clause-initial PPs in embedded clauses in Old English, together with the main verbs and their subjects. The data are analysed looking for syntactic and information–structural factors that may influence the positioning of the finite verb and the subject in these types of clauses.

6.1. PP Anaphoricity

In her study of clause-initial PPs in main clauses in the Old English *Orosius* and *Lives of Saints*, Dreschler (2015, p. 253) concludes that many of those clause-initial PPs are discourse-old and contain a link to the preceding discourse, functioning as local anchors, while less frequent but relevant patterns serve as frame-setting and contrast devices. In order to assess the situation in embedded clauses, the anaphoricity of clause-initial PPs, both with and without finite verb inversion was examined by analysing the elements immediately following the preposition. These elements are bare demonstratives (25), demonstratives which are part of a DP (26), bare pronouns (27) and pronouns which are part of a DP (28). Finally, PPs followed by a non-anaphoric element were also included (29)[2]. We need to bear in mind the fact that the Old English demonstrative could often function as a definite determiner instead of a true demonstrative.

(25)

forþon	*hit*	*gewis*	*is,*	*þæt*	***in***	***þam***	*beoþ*	*þa*
Because	it	certain	is	that	in	that	are	the
cwylmde	*æfter*	*þam*	*dome*					
tormented	after	the	judgement					

'. . .because it is certain that the bodily inclined will be tormented in that after judgement'
(GDPref_and_4_[C]:30.304.7.4517)

(26) *Is* *ðæt* *ec* *sæd* *þætte* ***in*** ***ðere*** ***stowe,*** ***þer*** ***hio***
Is that also said that in the place where they
ofslegne ***weran,*** *weolle* *an* *welle*
slain were sprang a well
'It is also said that in the place where they had been slain a well sprang forth'
(Bede_5:11.418.19.4207)

(27) *forþon þe* *hie* *wiston* *þæt* ***on*** ***hire*** *eardode* *se* *heofonlica* *cyning*
because they knew that in them dwelled the heavenly king
'. . .because they knew that the heavenly king dwelled inside of them'
(HomU_18_[BlHom_1]:11.148.135)

(28) *for ðan ðe* ***ðurh*** ***heora*** ***bodunge*** *is* *þes* *middaneard* *gebiged* *to* *ðam*
because through their preaching is this world turned to the
soðum *geleafan*
true faith
'. . .because this world is turned to the true faith through their preaching'
(+ACHom_II,_40:302.93.6882)

(29) *for þan ðe* ***on*** ***eastdæle*** *is* *þæs* *dæges* *angin*
because in east is the day's beginning
'. . .because the beginning of the day is in the east'
(+ACHom_I,_8:247.162.1522)

Table 2 below shows that 47.5% of the clause-initial PPs in embedded clauses with both word order types have no demonstratives or pronouns following the preposition, with mainly new material occupying that position. Concerning clauses with an anaphoric element following the preposition, those with a demonstrative within a DP represent 28% of the total. The data suggest that most initial PPs in embedded clauses have a frame-setting function. We cannot dismiss, however, a fair amount of examples which function as a link to the previous discourse, with anaphoric elements such as demonstratives and pronouns within the clause-initial PP.

Table 2. Anaphoricity of initial PPs in embedded clauses.

PP Anaphoricity	PP-S-V	PP-V-S	Both Types
P + Dem	11 (4.3%)	3 (0.9%)	14 (2.4%)
P + Dem DP	69 (27.1%)	95 (28.5%)	164 (27.9%)
P + Pro	16 (6.3%)	28 (8.4%)	44 (7.5%)
P + Pro DP	38 (14.9%)	25 (7.5%)	63 (10.7%)
P + No Dem/Pro	121 (47.5%)	182 (54.7%)	303 (51.5%)
Total	255 (100%)	333 (100%)	588 (100%)

Looking at the distribution of anaphoric elements within each type of word order, an even distribution can again be observed, with no sharp differences between clauses with and without verb inversion. The main conclusion that can be drawn from this fact is that the anaphoricity of the initial PP does not influence the positioning of the finite verb in the clause. Given that the anaphoricity of the clause-initial locative is one of the defining elements of locative inversion, this is an extremely surprising factor. Therefore, other clausal elements such as the subject or the finite verb itself should be examined in order to find any motivation for the positioning of the latter.

6.2. Subject Length

In relation to the status of subjects, the first variable that was analysed was their length in terms of number of words. The subjects of embedded clauses with clause-initial PPs were divided into five categories: subjects with one word, two words, three words, four words and more than four words. Table 3 below shows the distribution of subjects according to this categorisation in both types of word order.

Table 3. Subject length in embedded clauses with clause-initial PPs.

Subject Length	PP-S-V	PP-V-S	Both Types
1	112 (76.2%)	35 (23.8%)	147
2	78 (40.4%)	115 (59.6%)	193
3	41 (44.1%)	52 (55.9%)	93
4	8 (23.5%)	26 (76.5%)	34
4+	16 (13.2%)	105 (86.8%)	121
Total	255 (43.4%)	333 (56.6%)	588

There is a clear tendency for subjects in embedded PP-S-V clauses to fall under the category that we could define as 'short subjects', with almost half of them consisting of one word and 30% of two words. Subjects with four or more words with this type of word order are quite scarce in the data. In contrast, subjects in embedded PP-V-S clauses show a different distribution, with more than a third of them consisting of four or more words, falling under the category of 'long subjects'. If we pay attention to subordinate PP-V-S clauses with long subjects, more than 60% of those subjects contain a relative clause embedded within them, as in (30) below:

(30) *gif* ***on*** ***neawiste*** *gestrangiað* *þa* *þing* *þe* *Gode* *magon* *bringan*
if in presence strengthen the things that God can bring
selran *wæstm*
better fruit
'if in that presence the things that can bring profit to God strengthen'
(GD_2_[H]:3.108.32.1083)

Example (25) above displays a long subject with an embedded relative clause, i.e., *þa þing*, *þe Gode magon bringan selran wæstm.* The abundance of long subjects in embedded clauses with clause-initial PPs and verb inversion seems to indicate that there is a correlation between subject length and its position with regards to the verb. In this respect, López-Martínez (2019) points out that embedded clauses with clause-initial objects displaying an OVS word order could be the result of subject extraposition due to their informational status; it is not unlikely, then, for the same process to take place with clause-initial PPs.

This idea is supported by Warner's (2007) work on inversion in Late Middle English, where he demonstrates that longer subjects have a higher tendency to be inverted than shorter ones, with inversion taking place with subjects of four or more words in 97% of instances and with categorical inversion with five or more words (Warner 2007, p. 101). Warner (2007) relates this to the "Principle of End Weight" proposed by Quirk et al. (1972, p. 14.8), i.e., "the tendency to reserve the final position for the more complex parts of a clause or sentence".

Nevertheless, we must note that a PP-V-S word order with inversion does take place with short subjects as well, with one-word subjects being more numerous, for instance, than four-word subjects. We can relate this to the argument in Culicover and Levine (2001, p. 1) that there are two constructions within the phenomenon of locative inversion, i.e., light inversion and heavy inversion. In light inversion, which is restricted to unaccusatives, the postverbal subject can be "phonologically and structurally extremely simple", while heavy inversion is also possible with unergative or even transitive verbs, provided the subject is heavy. Thus, Culicover and Levine (2001, p. 3) assume that in light inversion, the subject is "in situ in VP", as in (31) below, while in heavy inversion, the subject "appears" in Spec-IP and then "postposes to the right of VP", as in (32).

(31) Light inversion: [IP PP I [VP V NP*subj* t . . .]]
(32) Heavy inversion: [IP PP [IP t'_{subj} I [$_{VP}$ t_{subj} V t_{PP} . . .] NP_{subj}]]
(Adapted from Culicover and Levine 2001, p. 3)

Although Section 6.4 will look at verb types in detail, we can observe that, of the 35 examples of embedded locative inversion with one-word subjects, only three appear with

lexical verbs, all of them unaccusative (*weaxan* 'grow' twice and *cuman* 'come'). The rest are examples of *be, have* and modals. These examples could be analysed as light inversion if we follow Culicover and Levine (2001). Looking at heavy subjects (four words or more), we find 131 instances of embedded locative inversion, with 30 lexical verbs. Although verbs like *weaxan* 'grow' and *cuman* 'come' also appear in this configuration, we can find unergative verbs such as *utgan* 'go out' or *ge-hyran* 'hear', and even transitive verbs like *ricsian* 'rule', thus being eligible to fall under the category of heavy inversion.

Unlike in main clauses with initial PPs and inversion, subject length seems to play a key role in inversion in subordinate clauses. This idea requires us to analyse the informational status of subjects in the dataset.

6.3. Subject Type

The last variable taken into account for the present analysis of subjects was subject type, based on the labels applied by the *YCOE* parsing, in order to assess their informational status. As observed in Table 4 below, the majority of subjects in embedded clauses with PP-S-V word order are pronominal, i.e., highly anaphoric material. Full NP subjects, quantified NPs and bare subjects are not abundant with this type of word order. Concerning an embedded PP-V-S word order with verb inversion, there are several key differences in the distribution of subject types. To begin with, we must note the abundance of NP subjects introduced by a demonstrative, which amount to more than a third of the data. Furthermore, there is a complete absence of pronominal subjects and a low number of bare demonstratives, which suggests that anaphoric elements do not motivate subject extraposition in this context. This also agrees with Webelhuth's (2011) fourth property of locative inversion in PDE (i.e., that the logical subject must not be an anaphoric pronoun).

Table 4. Subject type in embedded clauses with clause-initial PPs.

Subject Type	PP-S-V	PP-V-S	Both Types
Anchored NP	23 (76.7%)	7 (23.3%)	30
Bare	21 (42.9%)	28 (57.1%)	49
Dem	6 (66.7%)	3 (33.3%)	9
Dem NP	44 (26.7%)	121 (73.3%)	165
Full NP	29 (37.2%)	49 (62.8%)	78
Indep NP	8 (20.5%)	31 (79.5%)	39
Pro	64 (100%)	0 (0%)	64
Pro NP	2 (100%)	0 (0%)	2
Proper	12 (48%)	13 (52%)	25
Quant NP	27 (31.4%)	59 (68.6%)	86
Unknown	19 (46.3%)	22 (53.7%)	41
Total	255 (43.4%)	333 (56.6%)	588

Returning to the abundance of NP subjects introduced by a demonstrative, we need to consider the fact that the Old English demonstrative could function as a definite article, as is the case of *seo sunne* in (33) below:

(33) *for þan ðe* ***on*** ***nontide*** *asihð* *seo* *sunne*
because at ninth hour sets the sun
'because the sun sets at the ninth hour'
(+ACHom_II,_5:44.97.958)

A nominal element introduced by a definite article, like *seo sunne* ('the sun') in (20), does not need to be mentioned in the preceding discourse, and the information is identifiable by the hearer upon its first mention. Even though the subject in (20) is not long or heavy, its status as identifiable can motivate its extraposition, prompting the PP-V-S word order of the embedded clause.

6.4. Verb Type and Type of Clause

Returning to Webelhuth's (2011) principles of locative inversion in PDE concerning the main verb, it is necessary to remember that the finite verb must be impersonal and unaccusative. In order to establish whether this was the case in Old English embedded clauses with PP-V-S word order[3], finite verbs in the dataset were divided into four categories, namely 'be', 'have', modal verbs and lexical verbs. We must bear in mind that in the case of auxiliary 'be', 'have' and modal verbs, it is the unaccusative nature of the non-finite verbs in the dataset that meets the requirement for locative inversion. Table 5 below shows that the majority of finite verbs occupying the second position of embedded clauses with clause-initial PPs are forms of *be.* Lexical verbs are relatively common, while forms of *have* and modal verbs are scarce.

Table 5. Finite verb type in embedded clauses with clause-initial PPs and verb inversion.

Verb Type	PP-V-S
Be	235 (70.6%)
Have	2 (0.6%)
Modal	25 (7.5%)
Lexical	71 (21.3%)
Total	333 (100%)

In the case of forms of 'be', the most common in the dataset, they appear both as a copulative construction, as exemplified in (34) below, or as part of a passive one, as in (35). Copulative forms of 'be' and passives are some of the unaccusative contexts under which we would expect locative inversion to take place. We can relate this to Light's (2012) study on Early New High German, in which she argues that many copular clauses are existential/presentational, a context that strongly favours the extraposition of the subject. Light (2012, p. 176) suggests that subjects can be extraposed to express a narrow focus, as well as to provide "default accent on the subject", frequently in presentational contexts with subjects which are new to the discourse.

Concerning lexical verbs, which amount to over a fifth of the examples, the analysis of the dataset reveals that they are also unaccusative. Lexical verbs following the initial PP are usually unaccusative or convey a change in state, as is the case of *cymeð* in (36).

(34)

se	rihtwisa	is	heofen	gehaten	*for þan ðe*	***on***	***rihtwisum***
the	righteous	is	heaven	called	because	in	righteous
mannum	*is*	*Godes*	*wunung*				
men	is	God's	dwelling				

'the righteous is called heaven because God's dwelling is inside righteous men'

(+ACHom_I,_19:327.64.3674)

(35)

for þan þe	***þurh***	***þa***	***twa***	***þing***	*byð*	*þæt*	*eadige*	*lif*	*begeotan*
because	through	those	two	things	is	the	blessed	life	infused

'because the blessed life is infused through those two things'

(Alc_[Warn_35]:5.5)

(36)

þæt	***ofer***	***eow***	*cymeð*	*micel*	*storm*	*&*	*hreonis*
that	over	you	comes	great	storm	and	tempest

'that a great storm and a tempest will come over you'

(Bede_3:13.200.2.2024)

It can be concluded that, together with the informational status of subjects and subject length, the status of the finite verb as unaccusative does motivate the fronting of PPs in locative constructions in embedded clauses in the same way it does in main clauses. However, we must bear in mind Sasaki's (1998) notion that locative inversion in PDE is only possible in embedded clauses selected by bridge verbs. Van Kemenade (1997, p. 328)

defines bridge verbs as those that allow complementiser deletion, which can happen with 'that' in complement clauses. If Old English followed the bridge verb restriction Sasaki proposes for PDE, we would only find embedded locative inversion in complement clauses. We must therefore analyse the different types of clauses in which this structure appears in Old English. Figures 2 and 3 below show the distribution of clause type in embedded PP-S-V and PP-V-S clauses, respectively.

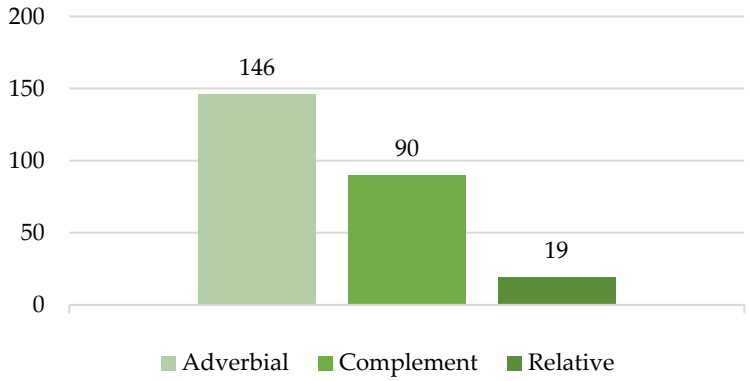


Figure 2. Embedded PP-S-V clause type.

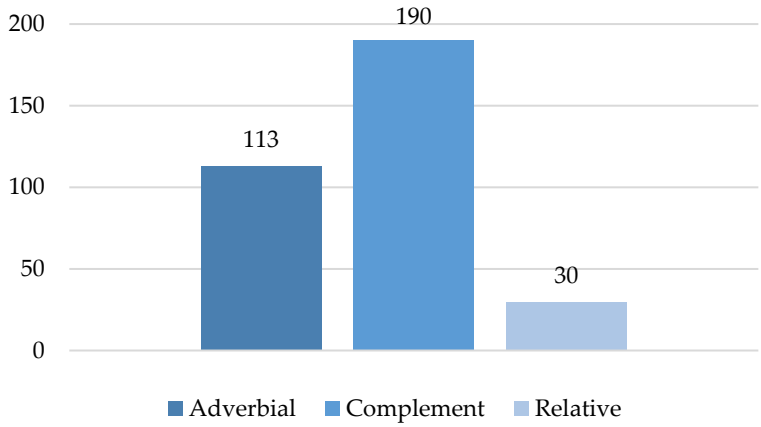


Figure 3. Embedded PP-V-S clause type.

While embedded clauses without finite verb inversion are adverbial in the majority of cases, most of those with PP-V-S word order are complement clauses, with a minority of relative clauses. Complement clauses in the dataset are indeed selected by bridge verbs such as 'say', 'think' or 'know'. Nevertheless, there are 113 instances of adverbial clauses, which are not introduced by a verb, as in (34), with *gehaten* preceding an adverbial clause introduced by *for þan ðe*. Thus, inversion in subordinate clauses with initial PPs is not restricted to complement clauses after bridge verbs in Old English.

7. Conclusions

The present paper was intended to assess the existence of a construction equal to locative inversion in embedded clauses in Old English. The analysis of the selected dataset proved that initial PPs can indeed appear in Old English embedded clauses with finite verb inversion, yielding PP-V-S word order. Numerous similarities were found with locative inversion constructions in PDE, such as the need for the verb to be intransitive and unaccusative and for the subject not to be an anaphoric pronoun. Looking at the anaphoricity of both initial PPs and late subjects, the relative familiarity constraint also applied, with most of the late subjects being new or retrievable information, always newer than the preposed element in the PP. Finally, the displaced effect is definitely at work in

this type of clause, which, in most cases, have a frame-setting function where the speaker has privileged access to the situation.

Even though anaphoricity does play a key role in this frame-setting function, it does not seem to be key in differentiating embedded clauses with clause-initial PPs with and without inversion. Unlike what happens with main clauses, where the type of verb was the key factor motivating inversion, subject length and type appears to play a much clearer role in subordinate clauses, with long and heavy subjects triggering finite verb inversion. Finally, the data suggest that locative inversion in embedded clauses is not restricted to those clauses selected by bridge verbs, as is the case in PDE, with a significant number of adverbial clauses displaying verb inversion after an initial PP. In conclusion, embedded clauses with clause initial PPs were a productive construction in Old English, resembling locative inversion in PDE. Most of the restrictions that have survived in PDE already applied, with the particular exception of that of bridge verbs. This could lead to future research concerning the implications for a syntactic model that would accommodate this type of construction.

Funding: This research was funded by Spain's Ministry of Science, Innovation, and Universities (MICINN), grant number PID2022-137233NB-I00.

Acknowledgments: I would like to thank Tara Struik for her technical support with the compilation of the dataset.

Conflicts of Interest: The author declares no conflict of interest.

Appendix A

Code used for the main query, using *Corpus Studio* (Komen 2009):

```
<TEI>
{
  (: Look for subclauses :)
  for search in //eTree[ru:matches(@Label, _subIP)]
  (: Look for PPs in initial position that are not empty :)
  let firstelement := search/child::eTree[ru:matches(@Label, _firstelement)
        and tb:PrecedingElement1(self::eTree)
        and not(exists(child::eLeaf[@Type="Star"]))][1]
(: Determine the element immediately following the object :)

  let sbj := search/child::eTree[ru:matches(@Label, _subject)
  and not(exists(child::eLeaf[@Type="Star"]))][1]
  let verb := search/child::eTree[ru:matches(@Label, _finiteverb)][1]
  (: Determine order of constituents :)
  let punct := search/child::eTree[ru:matches(@Label, ".|,")]
  let order := if (ru:relates(sbj, firstelement, "iFollows")) then "Obj-Sbj"
        else if (ru:relates(verb, firstelement, "iFollows")) then "Obj-Verb"
        else if ((ru:relates(punct, firstelement, "iFollows")) and (ru:relates(sbj, punct,
"iFollows"))) then "Obj-Sbj"
        else if ((ru:relates(punct, firstelement, "iFollows")) and (ru:relates(verb, punct,
"iFollows"))) then "Obj-Verb"
        else ()
  (: Create a database :)
  let db := tb:MakeaDatabase(firstelement, sbj, verb, order)
  (: Make sure this clause has a preposition and the right order :)
  where (
        exists(firstelement)
        and exists(sbj)
        and exists(order)
        )
```

```
    (: Return the main clause :)
    return ru:back(search, db, order)
}
</TEI>
```

Notes

1 This was part of a wider query, which included both fronted objects and PPs in subordinate clauses, with and without an inversion of the finite verb. Only data concerning fronted PPs were analysed for the present study.

2 These distinctions were based on Dreschler (2015).

3 The nature of the dataset did not allow for the obtainment of data concerning embedded clauses with fronted PPs without finite verb inversion.

References

Biberauer, Theresa, and Ans van Kemenade. 2011. Subject positions and information-structural diversification in the history of English. *Catalan Journal of Linguistics* 10: 17–69. [CrossRef]

Birner, Betty. 1996. *The Discourse Function of Inversion in English*. New York: Garland Publishing.

Bolinger, Dwight. 1977. *Meaning and Form*. London: Longman. [CrossRef]

Bowers, John. 1976. On surface structure grammatical relations and the structure-preserving hypothesis. *Linguistic Analysis* 2: 225–42.

Bresnan, Joan. 1994. Locative inversion and the architecture of Universal Grammar. *Language* 70: 72–131. [CrossRef]

Burzio, Luigi. 1986. *Italian Syntax: A Government-Binding Approach*. Dordrecht: Reidel. [CrossRef]

Chomsky, Noam. 2008. On phases. In *Foundational Issues in Linguistic Theory. Essays in Honor of Jean-Roger Vergnaud*. Edited by Robert Freidin, Carlos P. Otero and María Luisa Zubizarreta. Cambridge: MIT Press, pp. 133–66.

Coopmans, Peter. 1989. Where stylistic and syntactic processes meet: Locative inversion in English. *Language* 65: 728–51. [CrossRef]

Culicover, Peter W., and Robert D. Levine. 2001. Stylistic inversion in English: A reconsideration. *Natural Language and Linguistic Theory* 19: 283–310. [CrossRef]

Dreschler, Gea. 2015. Passives and the Loss of Verb Second: A Study of Syntactic and Information-Structural Factors. Ph.D. thesis, LOT Publications, Utrecht, The Netherlands.

Drubig, Hans Bernhard. 1988. On the discourse function of subject verb inversion. In *Essays on the English Language and Applied Linguistics on the Occasion of Gerhard Nickel's 60th Birthday*. Edited by Josef Klegraf and Dietrich Nehls. Heidelberg: Groos, pp. 83–95. [CrossRef]

Fischer, Olga, Ans van Kemenade, Willem Koopman, and Wim van der Wurff. 2000. The Verb-Second constraint and its loss. In *The Syntax of Early English (Cambridge Syntax Guides)*. Cambridge: Cambridge University Press, pp. 104–37. [CrossRef]

Haeberli, Eric. 2001. Speculations on the syntax of subordinate clauses in Old English. *Reading Working Papers in Linguistics* 5: 201–29.

Hale, Kenneth, and Jay Keyser. 2002. *Prolegomenon to a Theory of Argument Structure*. Linguistic Inquiry Monograph. Cambridge: MIT Press, vol. 39. [CrossRef]

Hoekstra, Teun, and René Mulder. 1990. Unergatives as copular verbs; locational and existential predication. *The Linguistic Review* 7: 1–79. [CrossRef]

Kim, Jong-Bok. 2003. English Locative Inversion: A Constraint-Based Approach. *Korean Journal of Linguistics* 28: 207–35.

Komen, Erwin. 2009. *Corpus Studio*. Nijmegen: Radboud University Nijmegen.

Komen, Erwin. 2012. *Cesax*. Nijmegen: Radboud University Nijmegen.

Kroch, Anthony, Ann Taylor, and Donald Ringe. 2001. The Middle English Verb-Second Constraint: A Case Study in Language Contact and Language Change. In *Textual Parameters in Older Languages*. Edited by Susan Hering, Peter van Reenen and Lene Schøsler. Philadelphia: John Benjamins Publishing Company, pp. 353–91.

Levin, Beth, and Malka Rappaport Hovav. 1995. Unaccusativity. *International Journal of Lexicography* 10: 258–64.

Light, Caitlin. 2012. The Information Structure of Subject Extraposition in Early New HighGerman. *UPenn Working Papers in Linguistics* 18: 169–77.

López-Martínez, Sergio. 2019. Embedded Constituent Fronting in Old English: Syntactic and Information Structural Implications. Ph.D. thesis, University of Oviedo, Oviedo, Spain.

Newmeyer, Frederick J. 1987. Presentational there-insertion and the notions root transformation and stylistic rule. Paper presented at 23rd Regional Meeting of the Chicago Linguistic Society, Chicago, IL, USA; pp. 295–308.

Ojea, Ana. 2020. Syntax meets discourse: Locative and deictic (directional) inversion in English. *Poznań Studies in Contemporary Linguistics* 56: 251–76. [CrossRef]

Pintzuk, Susan. 1991. Phrase Structures in Competition: Variation and Change in Old English Word Order. Ph.D. dissertation, University of Pennsylvania, Philadelphia, PA, USA.

Pintzuk, Susan. 1993. Verb seconding in Old English: Verb movement to Infl. *The Linguistic Review* 10: 5–35. [CrossRef]

Postal, Paul M. 1977. About a nonargument for raising. *Linguistic Inquiry* 8: 141–54.

Postal, Paul M. 2004. A paradox in English syntax. In *Skeptical Linguistic Essays*. Oxford: Oxford University Press.

Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech, and Jan Svartvik. 1972. *A Grammar of Contemporary English*. London: Longman.

Rizzi, Luigi, and Ur Shlonsky. 2006. Satisfying the subject criterion by a non subject: English locative inversion and heavy NP shift. In *Phases of Interpretation*. Edited by Mara Frascarelli. Berlin: Mouton de Gruyter, pp. 341–61. [CrossRef]

Rochemont, Michael, and Peter Culicover. 1990. *English Focus Constructions and the Theory of Grammar*. Cambridge: Cambridge University Press.

Salvesen, Christine M., and George Walkden. 2017. Diagnosing embedded V2 in Old English and Old French. In *Micro-Change and Macro-Change in Diachronic Syntax*. (Oxford Studies in Diachronic and Historical Linguistics). Edited by Robert Truswell and Éric Mathieu. Oxford: Oxford University Press, pp. 168–81.

Sasaki, Jun. 1998. On the Embedded and Matrix Locative Inversion. *Studies in European and American Culture* 5: 53–67.

Taylor, Ann, and Susan Pintzuk. 2012. The Effect of Information Structure on Object Position in Old English: A Pilot Study. In *Information Structure and Syntactic Change in the History of English*. Edited by Meurman-Solin Anneli, Maria Jose Lopez-Couso and Bettelou Los. Oxford: Oxford University Press, pp. 47–65. [CrossRef]

Taylor, Ann, Anthony Warner, Susan Pintzuk, and Frank Beths. 2003. *The York-Toronto-Helsinki Parsed Corpus of Old English Prose*. York: University of York, Electronic texts and manuals available from the Oxford Text Archive. [CrossRef]

Travis, Lisa de Mena. 1984. Parameters and Effects of Word Order Variation. Ph.D. dissertation, MIT, Cambridge, MA, USA.

Travis, Lisa de Mena. 1991. Parameters of phrase structure and verb-second phenomena. In *Principles and Parameters in Comparative Grammar*. Edited by Robert Freidin. Cambridge: MIT Press, pp. 339–64.

Van Kemenade, Ans, and Bettelou Los. 2009. Discourse Adverbs and Clausal Syntax in Old and Middle English. In *The Handbook of the History of English*. Edited by Ans van Kemenade and Bettelou Los. Sussex: Wiley-Blackwell, pp. 224–48.

Van Kemenade, Ans, and Tanja Milićev. 2012. Syntax and discourse in Old English and Middle English word order. In *Grammatical Change: Origins, Nature, Outcomes: Proceedings of DIGS VIII*. Edited by Diane Jonas and Stephen Anderson. Oxford: Oxford University Press, pp. 237–55.

Van Kemenade, Ans. 1997. V2 and Embedded Topicalization in Old and Middle English. In *Parameters of Morphosyntactic Change*. Edited by Ans van Kemenade and Nigel Vincent. Cambridge: Cambridge University Press, pp. 326–51. [CrossRef]

Warner, Anthony. 2007. Parameters of Variation between Verb-Subject and Subject-Verb Order in Late Middle English. *English Language and Linguistics* 11: 81–111. [CrossRef]

Webelhuth, Gert. 2011. Motivating non-canonicality in Construction Grammar: The case of locative inversion. *Cognitive Linguistics* 22: 81–105. [CrossRef]

Zwart, C. Jan-Wouter. 1991. Clitics in Dutch: Evidence for the position of INFL. *Groninger Arbeiten zur germanistischen Linguistik* 33: 71–92.

Zwart, C. Jan-Wouter. 1993. Dutch Syntax: A Minimalist Approach. Ph.D. dissertation, University of Groningen, Groningen, The Netherlands.

languages

MDPI

Article

Agreement and Information Structure in Spanish PRO$_{[PL]}$ *with*-DP †

María Mare [1,2]

[1] Faculty of Humanities, Universidad Nacional del Comahue, Neuquén Q8300, Argentina; maria.mare@fahu.uncoma.edu.ar
[2] Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), General Roca Q8332, Argentina
† To Manuel Leonetti.

Abstract: This paper aims to revisit a construction found in some Spanish varieties which refers to a set constituted by a singular referent and an annex introduced by the item *con* 'with': PRO(noun)$_{[PL]}$ *with*-DP. This construction triggers plural agreement and can be doubled by a plural pronoun, indicating that the annex is included in the set to which verbal agreement and the plural pronoun refer. For example, *Nosotros con Juan viajamos ayer* (literally, 'We with Juan travelled.1PL yesterday') means 'Juan and I travelled yesterday'. We explore the Spanish PRO$_{[PL]}$ *with*-DP, taking into account its discursive properties together with the syntactic requirements involved in the agreement patterns. In fact, although the two individuals denoted by this construction are involved as equal participants in the event, they have a different discursive status: one of them introduces new information, while the other refers to the immediate communicative situation. If some notions regarding information structure can be coded by binary features such as [+/−anaphor] and [+/−contrast], it is possible to find plurality triggered by the opposite combination of features within the same syntactic object. PRO$_{[PL]}$ *with*-DP is a possibility that the lexicons of some languages offer.

Keywords: pronouns; agreement; comitative constructions; Accessibility Theory; discourse anaphors

Citation: Mare, María. 2024. Agreement and Information Structure in Spanish PRO$_{[PL]}$ *with*-DP. *Languages* 9: 58. https://doi.org/10.3390/languages9020058

Academic Editor: Ana Ojea

Received: 30 November 2023
Revised: 30 January 2024
Accepted: 4 February 2024
Published: 6 February 2024

1. Introduction

Plural-person information opens an interesting range of options regarding the interpretation of referents at a discursive level. First-person plural, for instance, clearly refers to the speaker, but some common ground is necessary to complete the rest of the reference involved in the pronoun (1PL = 1SG + someone else). Interestingly, languages show different options to explicitly mention the 'rest of the reference' when it constitutes discourse-relevant information. The alternative this paper focuses on is found in different languages and, ever since Schwartz (1988), it has been known as the Plural Pronoun Construction (PPC). The main characteristic of the PPC is that the plurality shown by verbal agreement includes a singular referent and the argument introduced by a (not always covert) preposition-like item: a comitative item. For instance, *nous* 'we' and *avons* (AUX.1PL) in (1a) have a dual number reading, which includes the speaker and the argument introduced by *avec* 'with' (*mon frère* 'my brother'). All the examples in (1) have the same dual interpretation of plural morphology. The crosslinguistic differences lie in the possibility of dropping the plural pronoun (e.g., Catalan and Spanish vs. French) as well as in the presence of a covert comitative item (e.g., Icelandic)[1].

(1) a. Nous l' avons fait avec mon frère
we it did.1PL with my brother
'I did it with my brother' [French] (Rigau 1990, p. 366)

b. Amb la Maria us vau prometre pel febrer
with the Maria CL.1PL engaged.1PL on February
'I got engaged to Maria on February' [Catalan] (Rigau 1989, p. 203)

c. Con mi mujer nos casamos en abril
with my wife CL.1PL marry.1PL in April
'My wife and I got married in April' [Spanish] (Mare 2015, p. 275)

d. Me mentiin Annan kanssa kaupunkiin
we go.1PL Anna.GEN with town.ILL
'Anna and I went to the town' [Finnish] (Holmberg and Kurki 2019, p. 244)

e. Við Ólafur fórum.
we Olaf.NOM went.1PL
'Olaf and I went/left' [Icelandic] (Sigurðsson and Wood 2020, p. 3)

Following Holmberg and Kurki's (2019) remarks on the terminology used to describe the data, the constructions represented in (1) will be named *PRO(noun)$_{[PL]}$ with-DP* from now on. As has been widely discussed, these data indicate that the comitative item appears to be the element that triggers plural morphology, as if it were a Boolean phrase (see Lakoff and Peters' seminal 1969 work; Kayne 1994; Stassen 2000, among others). However, constructions in (1) differ not only from general coordination, but also from the Comitative Coordination (CC) found, for example, in Russian in *Maša s Dašej verjat v boga*, 'Maša and Dašej believe in God' (lit., *Maša with Dašej believe in God*, Feldman 2002, p. 43). First, in CC, the pronoun is singular when pronounced (DP$_{[-PL]}$ *with* DP ... verb$_{[+PL]}$) and, in many languages, it only involves two non-pronominal DPs (see Feldman 2002, pp. 42–43). The second difference is that neither of the DPs related by the comitative item in the CC can be moved out of the phrase (*DP ... verb$_{[+PL]}$ *with* DP/ **With* DP ... DP$_{[-PL]}$ ... verb$_{[+PL]}$). The relevance of pointing out these distinctions is that, in any case, the examples in (1) can be analyzed as instances of appositive constructions in which the *with*-DP phrase is an apposition of the plural pronoun (Mare 2012, for Spanish; Sigurðsson and Wood 2020, for Icelandic).

This paper aims to revisit the PRO(noun)$_{[PL]}$ *with*-DP in Spanish varieties taking into account its discursive properties together with the syntactic requirements involved in the observed agreement patterns. Since most of the previous analyses focus exclusively on the most frequent pattern (the 1PL exclusive reading), our goal is to develop a proposal that also captures the patterns that do not follow a person hierarchy but different discursive properties. Accordingly, we propose that this agreement is the result of a complex DP (as argued by Mare 2015) that refers to individuals with different informational statuses. We argue that there is a tight relation between the difference regarding information status and plural pronouns. In brief, there would be two mechanisms that trigger plural number: one of them operates with features related to person information, for instance, the combination of [−Participant]/[+Author] (Halle 1997), while the other involves oppositions in terms of anaphoricity. In a late insertion model like the one adopted in these pages, the externalization of the element that introduces the DP interpreted as 'the rest of the reference', i.e., the comitative item, follows from the syntactic structure transferred and the lexical items compatible with this structure. This means that if a variety does not have a lexical item for this structure, the derivation is ruled out.

The paper is organized as follows. Section 2 provides a general description of PRO$_{[PL]}$ *with*-DP in Spanish varieties and compares it with general comitative constructions (GCCs). Section 3 discusses some previous proposals regarding the PRO$_{[PL]}$ *with*-DP structure and presents an analysis for Spanish PRO$_{[PL]}$ *with*-DP. Section 4 focuses on information structure and the way in which this strategy works depending on the accessibility of the referents. Finally, we systematize the main conclusions drawn from this study.

2. On Spanish Comitative Constructions

Comitative constructions share some properties with plural DPs ($DP_{[PL]}$) and coordination ($DP_{[\&]}$) (see Conti Jiménez 2005, p. 297). This means that, if one of these entities refers to a human individual, for instance, then the other members involved will therefore be interpreted as having human properties, such as volition. The second similarity is closely related to the first: it is concerned with the fact that all the individuals denoted by these constructions relate to the event in a similar way in terms of their theta roles. Accordingly, there are some clear contexts in which these constructions are in complementary distribution, as shown below by the symmetric predicate *convivir* 'live together'.

(2)	a.	Mis	amigos		conviven.		[DP_{PL}]
		My	friends		live.together.3PL		
		'My friends live together.'					
	b.	Juan	y	Pedro	conviven.		[DP coordination]
		Juan	and	Pedro	live.together.3PL		
		'Juan and Peter live together.'					
	c.	Con	Pedro		convivimos.		[$PRO_{[PL]}$ *with*-DP]
		With	Pedro		live.together.1PL		
		'Peter and I live together.'					
	d.	Juan	convive		con	Pedro.	[GCC]
		Juan	lives.together		with	Pedro	
		'Juan lives together with Pedro.'					

In this section, we will discuss some relevant data that will hopefully provide us with a thorough description of the $PRO_{[PL]}$ *with*-DP construction in Spanish varieties. We will apply a number of syntactic and semantic diagnostics to distinguish between $PRO_{[PL]}$ *with*-DP (2c) and general comitative constructions (GCCs) like (2d). In Spanish, as well as in many other languages, both constructions present the same item: *con* 'with'. As will be shown, this comparison feeds Mare's (2012) hypothesis that $PRO_{[PL]}$ *with*-DP behaves as $DP_{[PL]}$, while GCCs present restrictions regarding the projections involved.

Most of the examples in this paper belong to varieties from Argentina and Chile. This is important to mention, because although the $PRO_{[PL]}$ *with*-DP construction is widespread, varieties differ in terms of the predicates with which it is combined and the person features involved. Furthermore, there are important differences in productivity, which, as we will show in Section 4, are related to communicative factors, rather than to grammatical properties.

2.1. Comparing PRO[PL] with-DP with GCCs

As has been pointed out in the literature (Mare 2012, 2013; Mare and Pato 2017), although $PRO_{[PL]}$ *with*-DP and GCCs can refer to the same number of participants involved in the event, there are some interesting differences between them. When it comes to verbal morphology, a GCC does not affect verbal agreement, regardless of its obligatory nature in terms of argument structure (3a). By contrast, in $PRO_{PL]}$ *with*-DP, the verb presents plural number information, despite the fact that the reference of the argument not introduced by *con* 'with' is a single entity ([-PL]). As can be observed, both the examples in (3) and (4) can be translated in the same way in the varieties under discussion, that is, the speaker with someone else (dual interpretation).

(3)	a.	Estoy	conversando	con	Angélica.	[GCC]
		be.1SG	speaking	with	Angélica	
		'I'm speaking with Angélica.'				
	b.	Estoy	escribiendo un trabajo	con	Angélica.	[GCC]
		be.1SG	writing a paper	with	Angélica	
		'I'm writing a paper with Angélica.'				

(4) a. Con Angélica estamos conversando. [PRO$_{[PL]}$ *with*-DP]
with Angélica be.1PL speaking
'I'm speaking with Angélica.'

b. Con Angélica estamos escribiendo un trabajo. [PRO$_{[PL]}$ *with*-DP]
with Angélica be.1PL writing a paper
'I'm writing a paper with Angélica.'

Beyond verbal agreement, the examples above display contrasts in constituent order. For example, *Con Angélica estoy conversando* 'I'm speaking with Angélica' and *Con Angélica estoy escribiendo un trabajo* 'I'm writing a paper with Angélica' are discourse-marked in comparison to the order in (3). In fact, one of the differences between the two comitatives regards information structure: while in the GCC, the unmarked constituent order is Verb-(Direct Object)- *with*-DP, in the PRO$_{[PL]}$ *with*-DP, the unmarked order presents the *with*-DP phrase on the left of the sentence. As will be shown in the following paragraphs, there are other diagnostics that allow us to go deeper in our understanding of the contrasts related to discursive aspects.

Mare (2012) presents specific diagnostics to distinguish PRO$_{[PL]}$ *with*-DP from GCCs. The first one is the combination of these two comitative constructions with other phrases headed by the item *con* 'with'. Pascual Pou (1999) observes that comitative and instrument phrases, which are both headed by *con* in Spanish, occupy the same position in the structure and, consequently, cannot appear together in the same sentence. We hypothesize that this position is related to the head Voice, i.e., the head that introduces agents in the syntactic structure (see Section 3). Following Pascual Pou's observations, Mare (2012) analyzes the combination of both comitative constructions not only with *con*-phrases introducing an instrument but also with *con*-phrases introducing company. As can be observed in the examples below, the differences are straightforward when maintaining the unmarked order for GCCs and PRO$_{[PL]}$ *with*-DP[2].

(5) Comitative + Instrument

a. ??Trabajé con Andrea con la computadora. [GCC]
work.1SG with Andrea with the computer
Lit., 'I'm working with Angélica with the computer.'

b. Con Andrea trabajamos con la computadora. [PRO$_{[PL]}$ *with*-DP]
with Andrea work.1PL with the computer
'Andrea and I are working with the computer.'

(6) Comitative + Company

a. ??Bailé toda la noche con Andrea con Severino. [GCC]
danced.1SG all the night with Andrea with Severino
Lit., 'I danced the whole night with Andrea with Severino.'

b. Con Andrea bailamos toda la noche con Severino. [PRO$_{[PL]}$ *with*-DP]
with Andrea danced.1PL all the night with Severino
'Andrea and I have danced with Severino the whole night.'

Changes in the order of the constituents can affect information structure in the PRO$_{[PL]}$ *with*-DP, whereas in the case of GCCs, these changes do not improve the results. Briefly, this contrast seems to feed the hypothesis that PRO$_{[PL]}$ *with*-DP and GCCs are associated with different parts of the syntactic structure.

A second distinction regards contrastive focus, i.e., a structure that opens the evocation of alternatives and resolves it through the pronounced option. When this evocation is not resolved, a *wh*-item is found. Interestingly, when the *with*-DP constituent is focalized, the dual interpretation of PRO$_{[PL]}$ *with*-DP is lost and the recovered referent is not a single entity but a plural one: we and Severino in (7a); you$_{[PL]}$ and someone else in (7b), capital letters representing the focalized item in (7a).

(7) a. #Con SEVERINO bailamos toda la noche.
with SEVERINO danced.1PL all the night
'We danced the whole night with Severino.'
b. #¿Con quién bailaron toda la noche?
with whom danced.2PL all the night
'Who did you dance the whole night with?'

Something similar happens with cleft sentences. When the VP is focalized, the dual interpretation of $PRO_{[PL]}$ *with*-DP is also lost.

(8) #Fue bailar con Severino lo que hicimos toda la noche.
was dance.INF with Severino the what did.1PL all the night
'It was dancing with Severino what we did the whole night.'

This diagnostic is relevant, because it feeds the hypothesis that in the constructions under study the *with*-DP phrase is part of a complex subject DP. Consequently, it cannot be focalized with the VP, as expected from any subject DP (9).

(9) *Fue bailar Juan lo que hizo toda la noche.
was dance.INF Juan the what did.3SG all the night

Another diagnostic that sheds some light on the behavior of $PRO_{[PL]}$ *with*-DP is found in control constructions. As is well known, these structures involve non-finite contexts in which the potential subject is unpronounced and controlled by one of the arguments in the main clause: I_i *want to* $\emptyset_i$ *see you tomorrow*. This null subject has been represented by PRO since Chomsky (1981), and its properties and distribution have led to the assumption that it is ungoverned according to the binding conditions of being both [+pronominal] and [+anaphoric]. The goal of this diagnostic is to define whether it is possible to obtain the dual interpretation of the 1PL in this context. In order to force the dual interpretation, we use a 1PL (10a) and a 1SG (10b) clitic as the controller and a pronominal construction for the subordinate verb. The result is that the dual reading is lost in both cases[3].

(10) a. #Juan nos recomendó comunicar -nos con Severino.
Juan CL.1PL recommended communicate.INF CL.1PL with Severino
'Juan recommended me to communicate with Peter.'
b. #Juan me recomendó comunicar -nos con Severino.
Juan CL.1SG recommended communicate.INF CL.1PL with Severino
'Juan recommended me to communicate with Peter.'

In sum, the dual number interpretation that characterizes $PRO_{[PL]}$ *with*-DP in Spanish seems to be closely intertwined with both verb–subject agreement and a specific discursive organization.

2.2. Holistic/Distributive Interpretation and Syntactic Functions

To continue with the comparison, let us revise some semantic characteristics of the structures under scrutiny. A relevant distinction refers to the combination of comitatives with different kinds of predicates. As Rigau (1989) explicitly remarks, across languages, comitative constructions contribute to a holistic interpretation of the predicate. This means that the argument introduced by the comitative does not trigger a plural event interpretation or, in other words, a distributive interpretation. For instance, in (11), the event of *working in Neuquén* is only interpreted as a singular event in (11a) as well as in (11b). Conversely, in (11c), it has a distributive reading (*Tom works in Neuquén and Jemmy works in Neuquén*).

(11) a. Tom works in Neuquén.
b. Tom works in Neuquén with Jemmy.
c. Tom and Jemmy work in Neuquén.

As Rigau observes, comitative constructions are incompatible with inherently distributive predicates, such as *know*, *understand*, *be a fan of*, etc. Interestingly, GCCs (12)

and PRO$_{[PL]}$ *with*-DP (13) present differences in their behavior in relation to this kind of predicates, at least in most of the Spanish varieties in which PRO$_{[PL]}$ *with*-DP is attested[4].

(12) a. *Soy fanática de River con mi hermano.
be.1SG a River fan with my brother
Lit., 'I'm a River fan with my brother.'

b. *Se tocar la guitarra con mi hermano.
know.1SG play the guitar with my brother
Lit., 'I know how to play the guitar with my brother.'

(13) a. Con mi hermano somos fanáticos de River.
with my brother be.1PL a River fan
'My brother and I are River fans.'

b. Con mi hermano sabemos tocar la guitarra.
with my brother know.1SG play the guitar
'My brother and I know how to play the guitar.'

This contrast is crucial to understand that each type of comitative phrase is related to the other referent in a structurally different way. A GCC merges in a position in which it does not affect the semantic interpretation of the event. In contrast, PRO$_{[PL]}$ *with*-DP seems to be part of an argument that must be interpreted as plural and, consequently, it also triggers the distributive reading. Of course, as Conti Jiménez (2005) observes, there seems to be a natural holistic interpretation with not-necessarily distributive predicates such as *viajar* 'travel' (14a). However, this interpretation appears to be part of an inference that also applies to coordination (14b) and plural DPs (14c), and not the result of syntactic compositionality.

(14) a. Con Juan viajamos en el verano. Él fue al mar y yo a la cordillera.
with Juan traveled.1PL in the summer He went to the sea and I went to the mountains
'Juan and I traveled during the summer. He went to the sea and I went to the mountains.'

b. Juan y yo viajamos en el verano. Él fue al mar y yo a la cordillera.
Juan and I traveled.1PL in the summer He went to the sea and I went to the mountains
'Juan and I traveled during the summer. He went to the sea and I went to the mountains.'

c. Los docentes viajamos en el verano. Él fue al mar y yo a la cordillera.
the teachers traveled.1PL in the summer He went to the sea and I went to the mountains
'We teachers traveled during the summer. He went to the sea and I went to the mountains.'

Another source of empirical data that supports this distinction is provided by quirky subjects, which—as is well known—are arguments of distributive reading predicates in Spanish. Again, in contrast with GCCs, PRO$_{[PL]}$ *with*-DP can freely occur in these constructions, at least in Argentinean and Chilean Spanish varieties[5].

(15) a. ??Con Ana me interesan esos libros. [GCC]
with Ana CL.1SG interest.3PL these books
Lit., 'I'm interested in these books with Ana.'

b. Con Ana nos interesan esos libros. [PRO$_{[PL]}$ *with*-DP]
with Ana CL.1PL interest.3PL these books
'Ana and I are interested in these books.'

This behavior is remarkable because quirky subjects present the dative case in Spanish and PRO$_{[PL]}$ *with*-DP does not fit well with syntactic functions related to accusative and dative case. This is an important difference between Spanish varieties and Slavic languages which allow PRO$_{[PL]}$ *with*-DP in all syntactic functions (Ionin and Matushansky 2002). Additionally, it is the main difference between PRO$_{[PL]}$ *with*-DP and DP$_{[PL]}$ or DP$_{[\&]}$. Grammaticality judgments on these examples in Spanish are not homogenous: in some cases, neither PRO$_{[PL]}$ *with*-DP (16) nor GCCs (17) are accepted, while in some others, the problem is that the plural marker is not interpreted as dual (18).

(16) a. ??Con Pedro nos saludó Juan. [**nos* = Pedro and me]
with Pedro CL.1PL greeted.3SG Juan
'Juan greeted Pedro and me.'
b. Con Pedro nos regaló un libro Juan. [**nos* = Pedro and me]
with Pedro CL.1PL gave.3SG a book Juan
'Juan gave Pedro and me a book.'

(17) a. Juan me saludó con Pedro. [*Pedro and me]
Juan CL.1SG greeted.3SG with Pedro
'Juan greeted Pedro and me.'
b. Juan me regaló un libro con Pedro. [*to Pedro and me]
with Pedro CL.1SG gave.3SG a book Juan
'Juan gave Pedro and me a book.'

(18) a. #Los militares nos interrogaron con mi hermano durante horas.
the military CL.1PL interrogated.3SG with mi brother for hours
'The military interrogated my brother and I for hours.'
b. #Los militares nos arrojaron con mi hermano a una celda.
the military CL.1PL threw.3SG with mi brother into a cell
'The military threw my brother and me into a cell.'
c. #Nos insultaron con mi hermano en las redes sociales.
CL.1PL insulted.3PL with my brother in the social media
'People insulted my brother and me in the social media.'

In spite of this restriction, when the passive construction is possible, $PRO_{[PL]}$ *with*-DP is allowed as subject (19), which suggests that the presence of $PRO_{[PL]}$ *with*-DP is strongly conditioned by verbal agreement, regardless of its thematic role interpretation.

(19) a. Con mi hermano fuimos interrogados durante horas.
with my brother were.1PL interrogated.PL for hours
'My brother and I were interrogated for hours.'
b. Con mi hermano fuimos arrojados a una celda.
with my brother were.1PL thrown.PL into a cell
'My brother and I were thrown into a cell.'
c. Con mi hermano fuimos insultados en las redes sociales.
with my brother were.1PL insulted in the social media
'My brother and I were insulted in the social media.'

2.3. Summary

All in all, there are clear differences between $PRO_{[PL]}$ *with*-DP and GCCs, as summarized in Table 1, which also accounts for the behavior of Spanish $DP_{[PL]}$ and $DP_{[\&]}$ in the contexts revised.

Table 1. Diagnostics for $PRO_{[PL]}$ *with*-DP syntactic and semantic behavior.

Diagnostics	$PRO_{[PL]}$ with DP	GComP	$DP_{[PL]}$	$DP_{[\&]}$
Verbal agreement	Yes	No	Yes	Yes
Combination with other *with*-DP phases	Yes	No	Yes	Yes
Holistic interpretation	Yes	Yes	Yes	Yes
Distributed interpretation	Yes	No	Yes	Yes
Quirky subjects	Yes	No	Yes	Yes
Accusative object	??	??	Yes	Yes
Dative object	??	??	Yes	Yes

As observed, the main contrast between $PRO_{[PL]}$ *with*-DP and other plural DPs regards their syntactic distribution. On the other hand, following previous studies in the literature, we mentioned some other contexts in which the dual interpretation of plural verbal agreement is lost (focalization of the *with*-DP and focalization of the VP). The next section revises the $PRO_{[PL]}$ *with*-DP structure.

3. The Internal Structure of PRO$_{[PL]}$ *with*-DP

Most of the diagnostics developed in Section 2 feed the hypothesis that PRO$_{[PL]}$ *with*-DP behaves as a plural DP and that the *with*-DP phrase is part of it (Feldman 2002; Vassilieva and Larson 2005; Vassilieva 2005). This fact would explain why the holistic and the distributive interpretation are both possible with PRO$_{[PL]}$ *with*-DP, although there is a comitative item that, in the general case, rejects the distributive reading. Moreover, if *with*-DP is part of a plural DP, agreement patterns follow without any further assumption: the verb agrees with the plural DP, not with a 'grammaticalized preposition' or a Boolean phrase that change its status throughout the derivation (Lakoff and Peters 1969; Kayne 1994).

In line with Ionin and Matushansky's (2002) analysis, we argue that the comitative item inside the Spanish PRO$_{[PL]}$ *with*-DP is a preposition, as long as the comitative item in the GCC is also analyzed as a preposition. This means that, beyond the proposed label, they behave in the same way, but each one appears in a different syntactic context. Stolz et al. (2006) note that the label *comitative* refers to the relationship between two entities belonging to the same entity class and participating in the event simultaneously. In harmony with these authors, Maslova (2007, p. 337) proposes the label 'participant set' to refer to the entities related by the comitative marker.

Inspired by these ideas, Mare (2020) argues that the comitative item introduces a DP but fails to assign a thematic role to it. The consequence of this failure is that the DP inherits the thematic interpretation from the projection to which the comitative item gets adjoined. In other words, in English, for instance, the interpretations for the item *with* and the DP introduced by it are obtained compositionally. If the DP has person features and the syntactic structure represents a transitive creation event (like *work*), it will be interpreted as part of the set referring to agents (*John works with Peter*). By contrast, if the DP refers to an inanimate entity and is introduced by *with* in the same syntactic structure, the interpretation obtained is that of an instrument, which is in some way linked to the agent (*John works with a hammer*).

This approach is also relevant to understand why GCCs cannot combine easily with other *with*-DP phrases. As mentioned above, both 'company' and 'instrument' are interpretations related to the agent and it is the head Voice that introduces the external arguments that are interpreted as agents and causers (Kratzer 1996). This means that *with*-DP$_{[COMPANY]}$ and *with*-DP$_{[INSTRUMENT]}$ occupy the same position in the structure, i.e., a projection adjoined to Voice. Conversely, PRO$_{[PL]}$ *with*-DP is compatible with other *with*-phrases because the DP introduced by the comitative is inside a DP and does not occupy the same position as other *with*-phrases.

In the following sections, we focus on the syntactic structure of PRO$_{[PL]}$ *with*-DP.

3.1. First-Person Plural as a Cue

Among plural pronouns, 1PL has the particularity of referring to a heterogeneous set of entities according to their participation in the communicative act. *We*, for instance, refers to the speaker and to another entity (or other entities) associated with it, which may be a part of the speech act (inclusive reading) or not (exclusive reading). As is well known, the general use of 1PL refers to the speaker and the group/entity associated with it. This description is very close to the notion of *associative plurality* developed by Daniel and Moravcsik (2013), who distinguish between associative and additive elements. In their words:

> An example of the additive plural is English *boys*. It is additive in the sense that it refers to a set where every member is a boy and thus the set is referentially homogeneous: every referent of the plural form is also a referent of the stem. In contrast, the associative plural designates a heterogeneous set.
>
> (Daniel and Moravcsik 2013, p. 1)

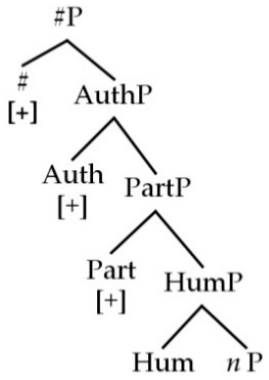
1PL is a clear example of an associative plural and, in fact, when Halle (1997, p. 129) decomposes person information in features such as Participant in Speech Event ([PART])

and Author in Speech Event ([AUTH]), he remarks that two apparently incompatible values can be combined in the syntax and that this combination triggers plural number. By way of illustration, [+PART] and [−PART] can coexist inside a DP insofar as this DP refers to a group formed by a hearer and someone else. Of course, the result of that is $DP_{[+PL]}$. The difference between a 'triggered' plural number and a 'free' plural number is hence tightly related to Daniel and Moravcsik's distinction.

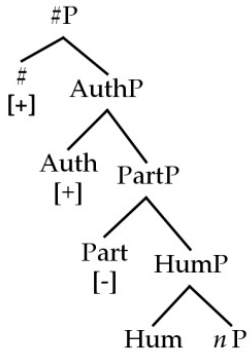
In the case of 1PL, following Halle's notion of a fourth person, we consider that plural number is triggered by the combination of [+AUTH] with either [+PART] or [−PART]. In both cases, the result is a referentially heterogeneous set, i.e., an associative plural.

In her approach to pronouns, Mare (2023) proposes a system in which Hum(an), Part and Auth project in the syntactic structure. HumP can merge with both Auth and Part. AuthP can be absent, while PartP can present a positive [+] or a negative [−] value. As was mentioned above, the feature [+PL] on # can be triggered by particular combinations of features. With that in mind, the structures for associative plural pronouns are formed in the syntax as schematized in (20), (21) and (22) (Mare 2023, p. 7)[6].

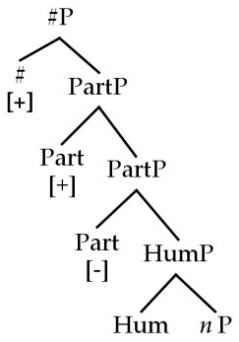
(20) 1PL inclusive



(21) 1PL exclusive



(22) 2PL exclusive



The previous structures shed light on the $PRO_{[PL]}$ *with*-DP construction. In some way, it seems to be a strategy that some varieties present in order to make explicit the part of the reference that is not linked to the communicative situation. In fact, in most of the examples, *with*-DP introduces a [−PART] referent which otherwise is lost. In Section 4, we show how this syntactic property relates to information structure.

3.2. The Syntactic Analysis of $PRO_{[PL]}$ with-DP

As mentioned in the introduction, the $PRO_{[PL]}$ *with*-DP construction is found in several languages. Feldman (2002) analyzes this construction in Russian and observes that speakers regard this construction as more natural when compared to ordinary coordination in which pronouns are involved. She also notes that the order of the elements related by the comitative item is restricted with respect to person hierarchy: the plural pronoun must be higher in the hierarchy (1 > 2 > 3) than the DP introduced by the comitative item. Feldman analyzes Russian instances of $PRO_{[PL]}$ *with*-DP as transitive plural pronouns that select the *with*-DP as the complement. To account for the fact that the denotation of the complement is included in the plural pronoun, Feldman assumes a pragmatic restriction according to which one element of a set comprises another element in its denotation (Feldman 2002, p. 60).

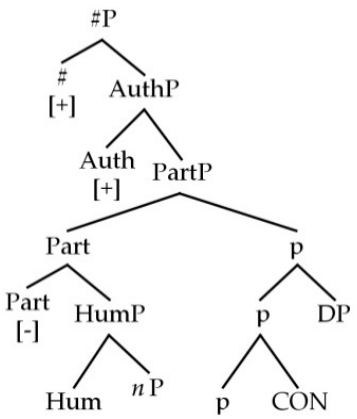
As has been mentioned, there are many proposals for this kind of construction in different languages, but Feldman's hypothesis highlights the close relation between $PRO_{[PL]}$ *with*-DP and the characteristics of plural pronouns outlined in Section 3.1. Kratzer (2009) proposes that associative elements present the feature [GROUP], which is independent of [+PL]. In particular, she notes that

> [I]n addition to speakers and addressees, contexts c may determine a function that assigns to selected individuals a not necessarily proper plurality of individuals that consist of the individuals themselves and their associates with respect to c.
>
> (Kratzer 2009, p. 224)

Following Kratzer's proposal, Mare (2012, 2013) introduced the feature [GROUP] in her analysis of comitative constructions in general as a property of the prepositional element that categorizes the root CON 'with', in line with the *Categorization Assumption* (Embick and Marantz 2008). This approach is interesting because it captures two properties of the constructions under study, i.e., their associative nature and the apparent lack of semantic content of the comitative item mentioned above. However, in a fine-grained analysis of pronouns, such as the one presented in Section 3.1, neither the feature [GROUP] nor a preposition categorizer are necessary to explain the $PRO_{[PL]}$ *with*-DP construction.

As discussed earlier, the interpretation of a referent with an associated group is obtained by combining features with different values, as in (20), (21) and (22). This kind of hypermarking triggers plural number. Furthermore, the idea that the comitative item introduces a DP but that the interpretation of this DP depends on its own referential properties and on the projection in which the *with*-DP merges can be codified without further assumptions. If arguments are introduced by relational projection and *con* 'with' does not add any semantics, the result of merging this projection—named *p*-CON, from now on—in the scope of PartP will give rise to the proper interpretation. The resulting structure is (23).

(23) $PRO_{[PL]}$ *with*-DP, 1PL exclusive



The structure in (23) presents the projections corresponding to the 1PL exclusive, which is the more frequent construction. The DP introduced by *p*-CON is interpreted according to the position *p*-CON merges with (i.e., PartP). Consequently, this DP makes the reference of [−PART] explicit. One advantage of this analysis is that it can account for the $PRO_{[PL]}$ *with*-DP construction in different languages. For instance, in a non-pro-drop language like French, ellipsis is impossible in this case and consequently the pronoun *nous* 'we' is going to externalize or lexicalize the projections of the main structure, while in Spanish varieties, they remain covert. On the other hand, in a language like Icelandic, in which $PRO_{[PL]}$ *with*-DP does not present something like a preposition to introduce the annex (*við Ólafur*, literally, 'we Olaf' for 'Olaf and I'), it could be the case that the relational projection *p* does not have a root as a complement. Of course, each language has its own characteristics regarding null elements and lexicalization; however, the basic structure of (23) could be extended to analyze data from other languages, beyond Spanish varieties[7].

Last but not least, according to our analysis, the three pronouns represented in (20–22) can be involved in the $PRO_{[PL]}$ *with*-DP construction. Nevertheless, there is a difference in productivity between the 1PL exclusive and the 2PL exclusive due to discursive factors. Accordingly, the less frequent $PRO_{[PL]}$ *with*-DP construction is the one that includes

[+Auth]/[−Part] (inclusive reading). In sum, the syntax of the varieties under analysis allows all the possibilities, and the difference in productivity is attributable to pragmatic factors. In the next section, we complete this panorama by making some remarks on information structure.

3.3. Summary

In this section, we reviewed some of the proposals that, as far as we understand, shed light on the data under study and offer some relevant answers to the behaviors exposed in Section 2. Accordingly, we focused on the analyses that understand the $PRO_{[PL]}$ *with*-DP as a plural DP (Feldman 2002; Vassilieva 2005). Moreover, the observations regarding the characteristics of the group denoted by this construction (Maslova 2007; Daniel and Moravcsik 2013) were relevant to a widespread pronominal element: the first-person plural pronoun. This element denotes a heterogeneous group with at least two different referents (Halle 1997). They differ in terms of their participation in the speech act. With these ideas in mind, we design a proposal that takes into account pronoun properties and the distinction between homogeneous plurality and heterogeneous plurality. As a result, we consider syntactic structures for associative plural pronouns (20 to 22) that can be found in any language, but in some of them, these structures allow for the introduction of a DP. The introducer can be an overt *p*, as in Spanish, or a covert *p*, as in Icelandic. The advantage of this proposal is that it allows the derivation of all the possible combinations of features that give rise to associative plurals without adding any particular theoretical assumptions. Of course, when discussing variation, it will be crucial to observe in which cases a DP can be introduced in each dialect, under the scope of an associative plural pronoun. In the next section, we revise the notion of heterogeneous plurals codified in syntax, in addition to considering information structure in order to account for other combinations of features.

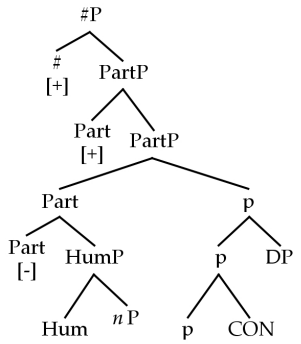
4. Information Structure and $PRO_{[PL]}$ *with*-DP

It is impossible to talk about $PRO_{[PL]}$ *with*-DP without revisiting the referential properties of plural pronouns. These elements have curious properties in terms of information structure also. As has been mentioned, 1PL refers to a group which is heterogeneous in terms of participation in the speech act. In the inclusive reading, 1PL refers to the author of the speech act (Auth) and the participant of the speech act (Part), while the exclusive reading involves the speaker and a non-participant. Although this last reading is not the only one available when talking about $PRO_{[PL]}$ *with*-DP, it is clearly the most frequent one. Something similar is found in relation to 2PL: it can refer to a homogeneous group of participants or to a group constituted by a participant and a non-participant. The latter option is exemplified in (24), where A is taking about holidays to B, who knows that Juan is A's partner.

(24) A: - Estoy de vacaciones.
'I'm on holiday.'
B: - ¿Se van a algún lado con Juan?
CL go.2PL to some place with Juan?
'Are you travelling somewhere with Juan?'

The corresponding structure for the $PRO_{[PL]}$ *with*-DP sequence in B's interaction is represented in our analysis by adding the *p*-CON to the pronominal construction in (22), as in (25).

(25) $PRO_{[PL]}$ *with*-DP, 2PL exclusive



Finally, it is essential to pay attention to 3PL, even though $PRO_{[PL]}$ *with*-DP is not found as frequently in the 3PL as in the 1PL and the 2PL. As is well known, 3PL is anaphoric by nature, so the heterogeneity involved in $PRO_{[PL]}$ *with*-DP cannot be understood in terms of participation in the speech act but in terms of topic orientation (Leonetti 2022). Let us observe the example in (26).

(26) - ¿Sabés algo de Paco?
'Any news from Paco?'
- Están entrenando con Ali para la Regata.
be.3PL training with Ali for the canoe race
'He is training for the canoe race with Ali.'

Neither of the referents for the 3PL found in the answer—i.e., Paco and Ali—are part of the speech act. Nevertheless, they present differences regarding information structure: the reference to Paco in the answer is retrieved from the question, while Ali is part of the new information.

The relevance of these data is that they change the way in which the relation between referents inside the $PRO_{[PL]}$ *with*-DP is analyzed. Following typological studies (Stassen 2000; Stolz et al. 2006, 2013, among many others), it is attractive to look for explanations in terms of person hierarchy, and the most frequent data feed this kind of approach: 1PL agreement is found when Auth is the unpronounced referent and the p-CON introduces the "annex". This approach correlates heterogeneity in the kind of participation in a speech act with person hierarchy: the syntactic structure of $PRO_{[PL]}$ *with*-DP implies a structural hierarchy in which the most marked participant is higher and the less marked participant is introduced by the comitative. Nevertheless, the data regarding 3PL lead us to suspect a strictly syntactic explanation, based on features, and invite us to look for more data, independently of its frequency.

Mare and Pato (2017, p. 85) mention a very interesting set of data in which the comitative introduces the speaker as new information because the unpronounced referent is recovered by an antecedent. In (27), the antecedent is *Antorcha Campesina*, a 3SG referent, and in (28), the antecedent is the group formed by *Quilco* and *Agiali*, both PART (2PL). Examples (27) and (28) are adapted from Mare and Pato.

(27) Quiero una relación cercana y de entendimiento con Antorcha Campesina;
'I want a close relationship and understanding with Antorcha Campesina'

con	-migo	nos	vamos	a entender bien.
with	-1SG	CL.1PL	AUX.1PL	to undertand well

'we will understand each other well'.
antorchacampesina.org.mx. 26/09/2015, Mexico

(28) –Oye –dijo Quilco a Agiali– vamos a coger cañas.
'-Hey- said Quilco to Agiali- let's go get some reeds.
–¿Y si nos ven? –objetó Agiali (. . .).
-What if they see us? -said Agiali (. . .).

–No hay nadie por ese lado.	Vamos	con	-migo
There's nobody on that side.	go.1PL	with	1SG
	- Let's go with me.		

Raza de bronce, Alcides Arguedas, 1919, Bolivia

In brief, following Leonetti's (2022) and Leonetti and Escandell Vidal's (2020) way of reasoning, we claim for a division of labor between syntax and pragmatics. Syntax must explain the abstract structure of the $PRO_{[PL]}$ *with*-DP, the mechanisms of agreement and the reasons why plural number is obtained, among other things. On the other hand, pragmatics should deal with referent identification, i.e., which is the most salient accessible antecedent in the discourse. Interestingly, this labor exceeds 3PL: 1PL and 2PL are also ambiguous between the dual reading ($PRO_{[PL]}$ *with*-DP) and the plural reading (GCC), but it is the context that is going to define the optimal antecedent for the referent accompanied by the *with*-DP phrase.

With this information in mind, let us remark on the extent to which syntax and pragmatics dialogue. As discussed in Section 3, some features' combinations trigger plural number. We focus on person features, but, as is well known, first and second person are dependent on the discursive context, so, accordingly, features like [Auth] and [Part] are necessarily related to pragmatic issues. Moreover, these features guarantee the heterogeneity needed for plural number: [Auth] needs to combine with other person features to obtain plurality and [Part] also opens this option.

Nevertheless, there are two facts that draw our attention: the first is the behavior of 3PL just mentioned and the examples in which the *p*-CON introduces a 1SG pronoun (27 and 28); the second refers to the diagnostics developed in Section 2, specifically, the data where some kind of focalization takes place and the dual reading is lost. Section 4.1 focuses on the difference between the two referents retrieved by $PRO_{[PL]}$ *with*-DP in terms of accessibility, while Section 4.2 discusses the cases in which the dual reading is lost.

4.1. Codifying Salience

The analysis developed in Section 3 moves the debate to the distinction between overt subject pronouns and null subjects. As proposed, $PRO_{[PL]}$ *with*-DP in Spanish is a peculiar kind of element which presents a null part and an overt part at the same time. The null part is recovered from the discourse situation, while the comitative introduces a referent that is otherwise ambiguous or hardly accessible. Moreover, $PRO_{[PL]}$ *with*-DP is compatible with overt plural pronouns, a fact that leads us to pay attention to the distribution of null subjects and the structure under study.

Some authors point out that null subjects are topic oriented (see Samek-Lodovici 1996, p. 46; Camacho 2013, p. 146) and that overt preverbal subjects must be licensed by topic antecedents. On the other hand, Leonetti (2022) notes that the relevant notion is not topicality but salience. In fact, the author proposes that, according to the data, a null subject may be linked to a non-topical antecedent if it is salient enough. According to the Accessibility Theory (Ariel 1990, 2001), human languages provide speakers with the means or mechanisms to encode the degree of accessibility of referents, i.e., complex systems of markers that indicate whether or not the mental representation of the referent being alluded to is immediately accessible to the listener. The less accessible a referent is, the more elaborate the strategy used to make it explicit is. Ariel (1990) puts forward the proposal that referential expressions have the function of accessibility markers that signal the addressee which is the mental representation of the referent. As mentioned, in 1PL pronouns, for instance, the accessible referent is the speaker, but the rest of the reference is not accessible enough unless it can be recovered from the previous context. Moreover, it

seems that the inclusive interpretation is the most accessible by default, as can be observed in the 1PL imperative construction: *¡Hagámoslo!* 'Let's do it'.

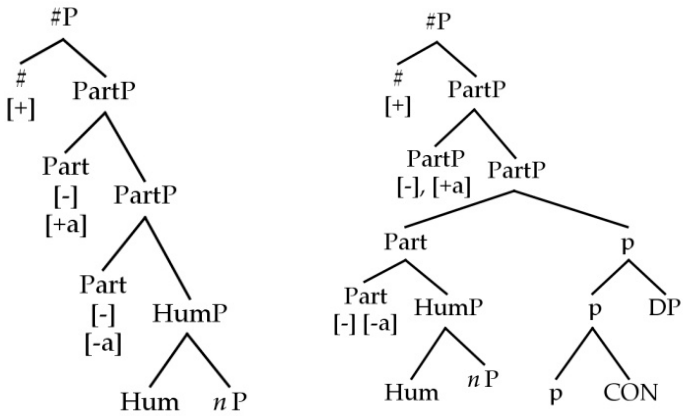
In terms of accessibility, $PRO_{[PL]}$ *with*-DP seems to be an intermediate option between a null plural pronoun and the coordination of DPs. The former is not informative enough, while the latter places the two coordinated DPs at the same informative level. The $PRO_{[PL]}$ *with*-DP strategy allows the recovery of a hierarchy of accessibility between the referents involved. In fact, inside this structure, the two poles of the scale proposed by Ariel (1990, p. 73; 2001, p. 21) can coexist: a null element and a full name.

The high frequency of $PRO_{[PL]}$ *with*-DP with 1PL agreement is expected for two reasons. First, this construction is related to informal interchanges, something already observed by (Kany [1945] 1970, p. 314), Camacho (1999, pp. 2669–70) and RAE-ASALE (2009: §33.7i). Second, the speaker is clearly an accessible referent. Beyond this, $PRO_{[PL]}$ *with*-DP can present different features and recover different referents: processing works insofar as the unpronounced referent is salient enough for the interlocutor to be able to retrieve it.

Inspired by López (2009), we propose that some of the properties mentioned above can be codified in syntax by using two features: [+/−a(naphor)] and [+/−c(ontrast)]. However, we follow Leonetti (2022) in considering that not all sentence topics, i.e., elements marked as [+a], are equated with left dislocated phrases. Accordingly, null subjects in general are going to be marked as [+a]—in fact, they are definite—and not anaphoric elements, as [−a]. On the other hand, new information is going to be related to the feature [+/−c]. As López (2009, p. 37) proposes, fronted focus constituents and *wh*-phrases are [+c], while regular focus is [−c] (and [−a], of course). In the light of these distinctions, $PRO_{[PL]}$ *with*-DP can be understood as a complex DP in which the difference—in terms of accessibility—between the two participants it refers to is codified by the combination of the features [+a] and [−a, −c]. The former ([+a]) is responsible for the topicality of one of the referents, while the latter set ([−a, −c]) marks the *with*-DP phrase. Interestingly, the fact that two opposite features coexist inside the same DP necessarily triggers plurality and codifies the interpretation of a heterogeneous group developed in Section 3.1.

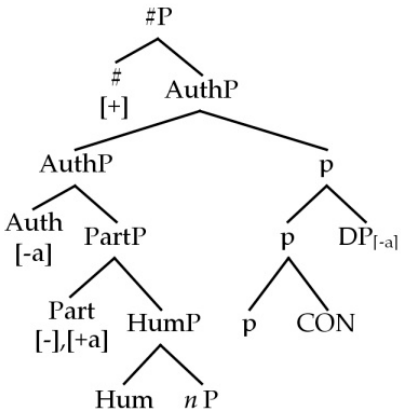
The structure in (29) represents the situation in which there is an associative group that refers to two entities that do not participate in the speech act (example 26). The configuration is similar to 2PL (exclusive), but the duplication of the head Part in (29) is due to the difference between the referents in terms of accessibility.

(29) a. 3PL- associative reading b. $PRO_{[PL]}$ *with*-DP, 3PL



Finally, the structure in (30) would represent the rare cases where the verb shows 1PL agreement but where the antecedent is not the speaker but some other referent (examples 27 and 28). The base is a 1PL configuration, but the features [+a] and [−a] are distributed differently from the most regular cases. This causes that *p*-CON merges in the structure in relation to AuthP, resulting in the expected introduction of a first-person pronoun.

(30)



In brief, the frequency of $PRO_{[PL]}$ *with*-DP that triggers 1PL agreement and where the recovered referent is the speaker is not due to a kind of person hierarchy but to discursive factors. Leaving this aside, the contribution of syntax in the inference process is restricted to features that allow us to explain the absence of phonological material for the salient referent, on the one hand, and the mandatory nature of plural number, on the other. In the next subsection, we will present some arguments in favor of a syntax that codifies this kind of information in some way.

In Section 2.1, we discussed a set of data in which the dual interpretation was lost and the comitative phrase cannot be part of the plural subject. The contexts involve focalization and *wh*-phrases. The relevant examples are repeated below.

(31) a. #Con SEVERINO bailamos toda la noche.
with SEVERINO danced.1PL all the night
'We danced the whole night with Severino.'

b. #¿Con quién bailaron toda la noche?
with whom danced.2PL all the night
'Who did you dance the whole night with?'

According to the hypothesis that the $PRO_{[PL]}$ *with*-DP construction is a partially null pronoun in which both [+a] and [−a] coexist, the result of focalization is expected and reinforces the proposal that the *with*-DP phrase is part of a complex DP. As López (2009) remarks, focus constituents and *wh*-phrases are [−a], but, most importantly, they are [+c], a feature connected with the left periphery. Even if we were to assume with López that [+a] constituents are also left dislocated, there is a clear incompatibility regarding the final locus in each case and, more obviously, regarding the materialization of the referent's information. The feature [+c] not only implies movement, but also phonological materialization, i.e., it is incompatible with null information. In brief, the coexistence of the features [+a] and [+c] in a DP seems to be impossible.

Furthermore, the construction under study can be doubled by an overt pronoun, as in (32).

(32) Nosotros nos casamos en abril con Ana.
we CL.1PL married.1PL in April with Ana
'Ana and I got married in April.'

The overt pronoun can present the features [−a, −c] and constitutes a proper antecedent for the [+a] feature of $PRO_{[PL]}$ *with*-DP. However, when the overt pronoun presents a contrastive reading, the dual interpretation (Ana and I) is lost, as shown in the translation of B's answer in (33).

(33) A: Juan y Julia se casaron en abril.
Juan and Julia CL.3PL married.3PL in April
'Juan and Julia got married in April.'
B: #NOSOTROS nos casamos en abril con Ana.
we CL.1PL married.1PL in April with Ana
'It was us who got married in April with my wife.'

The relevance of this observation is that the harmony among features when a pronoun is overt explains the $PRO_{[PL]}$ *with*-DP in non-pro-drop languages like French (remember (1a)): the obligatory overt subject is a plural pronoun which becomes more accessible owing to the presence of a structure with compatible discursive features.

The last point we would like to include here regards the occurrence of the overt plural pronoun. As has been mentioned, it does not have a contrastive reading. In addition to the constituent order in (34), the one represented in (35) is also possible. In both orders, the overt pronoun appears to the left of $PRO_{[PL]}$ *with*-DP. Remarkably, the relevant interpretation is lost when this order changes (32).

(34) Nosotros con Ana nos casamos en abril.
we with Ana CL.1PL married.1PL in April
'Ana and I got married in April.'

(35) #Con Ana nos casamos en abril nosotros.
with Ana CL.1PL married.1PL in April we
'My wife and we got married in April.'

Leonetti and Escandell Vidal (2020) note that strong pronouns as subjects are in competition with null subjects, because the former are used only when certain specific information-structure factors justify them. This observation involves preverbal as well as postverbal strong pronouns, and a factor that justifies them, beyond contrast or emphasis, is referent identification. We hypothesize that in cases like (32) and (34), the strong pronoun is a topic originally merged at the left periphery of the sentence and that in most contexts it is used to change the reference from 1SG to 1PL. This means that the subject in (32) and (34) is just $PRO_{[PL]}$ *with*-DP. Although this idea should be explored further, we believe that this change of reference is the key to understanding why the order in (35) blocks the dual number interpretation for 1PL.

The hypothesis developed here opens an interesting panorama for language variation. Syntax can combine features in different ways, but not all languages can lexicalize each combination. The Spanish varieties discussed in this paper lexicalize the combination [+a]/[−a] with a $PRO_{[PL]}$ *with*-DP, but this option is not spread through all Spanish varieties. Needless to say, further research on this perspective could offer an explanation for the restrictions observed in Section 2.2 regarding syntactic functions.

4.2. Summary

This section delved into $PRO_{[PL]}$ *with*-DP from an information-structure perspective. This construction has the same distribution as a null subject and, accordingly, it is unable to convey contrast. The covert referent is contextually determined, and the *with*-DP constitutes new information. These two properties are codified by the features [+a] and [−a], respectively, which allow us to codify in the syntax the accessibility proposed for the interpretation of the referents. Finally, we argued that the presence of these features in the syntactic domain explains the mandatory presence of plural number for the complex DP and the impossibility of maintaining the dual number reading when the *with*-DP is involved in a fronting focus or in a *wh*-constituent.

5. Final Remarks

The analyses of $PRO_{[PL]}$ *with*-DP, at least in Spanish varieties, cannot be complete if we do not take into consideration syntactic aspects as well as informative characteristics. The key feature is that this structure is the strategy some languages offer to materialize

an asymmetric informative relation between the two referents involved. This asymmetry triggers plural number because the syntax of $PRO_{[PL]}$ *with*-DP involves hypermarking: the combination [+Auth]/[−Part], [+Part]/[−Part] or [−Part] [+a]/[−Part] [−a]. Moreover, the possibilities found are not restricted to any kind of person hierarchy, as a syntactic property of languages, but to discursive patterns which affect the frequency of use for each possibility. As we have seen, a 1SG pronoun can be introduced as new information and so is the annex for an anaphoric 3SG, but the most frequent discursive situation is the opposite, where the most accessible referent is the speaker, and, consequently, it can be null.

In syntactic terms, we propose that *with*-DP is part of a pronominal structure in which plural number is triggered by feature hypermarking. The general interpretation of 1PL as an associative plural is a cue to understand the syntax of $PRO_{[PL]}$ *with*-DP and its referential properties. The dialogue between syntax and information structure offers the adequate frame to analyze the sequence under study.

A few words need to be said about variation in Spanish. As was pointed out throughout the paper, associative/heterogeneous plural pronouns can be found in all languages, but $PRO_{[PL]}$ *with*-DP is restricted to certain languages/varieties. As shown in (1), this property is not conditioned by being a pro-drop language or a non-pro-drop language. Our hypothesis is that the merging of a *with*-DP in the context of pronominal projections is not allowed in the grammar of many languages as a strategy for identifying a less salient referent. In fact, Spanish varieties differ in this aspect, because in some varieties the $PRO_{[PL]}$ *with*-DP is not observed and speakers do not make the dual interpretation when exposed to these data. At the opposite extreme, there are varieties in which the *p*-CON can merge in relation to AuthP, as discussed in Section 4.1.

This description could open relevant research regarding microvariation and containment: if a variety presents $PRO_{[PL]}$ *with*-DP, 3PL—the structure represented in (29b)—it also presents $PRO_{[PL]}$ *with*-DP, 2Pl and 1PL; if a variety allows *p*-CON in relation to AuthP, it also allows it in relation to PartP. All in all, the structures and features proposed in this paper are consistent with current studies of (micro)variation (see Bobaljik 2012 and the Nanosyntax literature inspired in his proposal). A final remark on variation concerns the distinction between syntactic possibility and pragmatic oddness. The lower frequency of a particular combination of features does not mean that this combination is avoided by the syntax. This is a point that is relevant to the way our analysis was conceived.

Of course, there are some aspects that deserve more research. The most remarkable one regards syntactic functions: $PRO_{[PL]}$ *with*-DP cannot appear as an accusative or a dative object. This is an important difference from the Slavic version of $PRO_{[PL]}$ *with*-DP, and it also contrasts with the fact that it can be the subject of passive constructions (i.e., it can be an internal argument) and materialize quirky subjects (dative). Probably, further research on information structure could give us some answers to this puzzle.

Funding: This research received no external founding.

Institutional Review Board Statement: Not applicable since the study did not involve human subjects.

Informed Consent Statement: Not applicable since the study did not involve human subjects.

Data Availability Statement: All the relevant data is available in the paper.

Acknowledgments: I am thankful to the anonymous reviewers whose expertise contributed to improve this manuscript. I would also like to thank Juanjo Arias for his invaluable help and Montse Batllori for encouraging me to write this paper.

Conflicts of Interest: The author declares no conflict of interest.

Notes

[1] Of course, all the examples in (1) are ambiguous between a dual and a plural reading of verbal agreement and of the plural pronoun in the cases in which it is materialized.

[2] One of the external reviewers disagrees with the judgments and observes that the oddity of the sentences in (5a) and (6a) may be due to some phonetic constraint related to the occurrence of two PPs with the same preposition in a row. However, this does not seem to be the case, since the same sequence is much better when the $PRO_{[PL]}$ *with*-DP is possible: *Trabajamos con Andrea con la computadora* 'Andrea and I worked with the computer'. The reviewer notes also that the sentences in (5a) and (6a) improve when something is added in between (*Trabajé con Andrea ayer con la computadora* 'Yesterday I worked with Andrea on the computer') or when the DP introduced by *con* is modified (*Corté la leña con Juan con un hacha muy afilada* 'I cut the firewood with Juan with a very sharp axe'). I agree with the reviewer in these judgments, but it is also true that the addition of some elements can improve ungrammatical or odd sentences for various reasons, including changes in the syntactic hierarchy. Regardless of these observations, in the case of (5b) and (6b), no change is needed to improve the sentence.

[3] In Section 4, we come back to this diagnostic, because the impossibility of maintaining the dual reading is also related to the properties of null subjects in control structures.

[4] Camacho (2000) remarks that $PRO_{[PL]}$ *with*-DP is not compatible with distributive predicates in Spanish. However, Mare (2012 and subsequent work) argues against extending Camacho's observation to all varieties of Spanish, as if it were a property of $PRO_{[PL]}$ *with*-DP. In fact, the data show a general dialectal extension $PRO_{[PL]}$ *with*-DP combined with distributive predicates. It is not difficult to find data via the web search: *con mi hermana somos fans de gossip girl* 'My sister and I are fans of Gossip girl' (https://twitter.com/shawnxito/status/1749454339431604539, Argentina, accessed on 29 January 2024).

[5] The following examples can be found on the Internet:

(i) *con mi hermana nos gusta experimentar en la cocina* [Argentina]
'My sister and I like to try new thing in the kitchen'
https://www.clarin.com/zonales/abuelo-invento-galleta-marinera-hoy-mantienen-panaderia-historica-conurbano_0_tFt8r3zeR.html (accessed on 30 January 2024)

(ii) con mi hermana nos gusta mucho la artesanía [Chile]
'My sister and I like to handcraft a lot'
https://cl.socialab.com/challenges/AcademiadeImpacto2021/idea/137341 (accessed on 30 January 2024)

[6] Interestingly, the overt 1PL pronoun *nosotros* and the 2PL pronoun *vosotros*—which are characteristic of some varieties—are morphologically complex. Eberenz (2000) points out that combinations of *nos* (1PL)/*vos* (2PL) and *otros* 'others' began to be used sporadically in the thirteenth and fourteenth centuries "to add a more specific content to the deictic meaning of the pronoun" (2000, p. 58). In fact, the combination of *vos* with *otros* 'others' or *todos* 'all' was employed, according to the author, to differentiate the contexts in which the 2PL pronoun referred to a sum of second persons (additive reading) from those in which it referred to a second person with an associated group (associative reading).

[7] Spanish varieties may present an overt pronoun (see the examples in Section 4.2 and below). In pro-drop languages, this overt pronoun is found under certain pragmatic conditions, whereas in non-pro-drop languages, such as French, the covert pronoun is obligatory. This is discussed in Section 4.2.

(i) Ellos pueden ver a sus familias. Nosotros con mi mujer no tenemos consulo [Argentina]
'They can see their families. My wife and I have no consolation'
https://www.lavoz.com.ar/sucesos/padre-de-baez-sosa-sobre-asesinos-lo-mataron-como-en-una-jauria-no-le-dieron-ni-una-oportuni/ (accessed on 26 January 2024)

(ii) Nosotros estamos muy felices con mi señora y mis seis niños [Colombia]
'My wife, my six kids and I are very happy'
https://www.elcolombiano.com/historico/uraba_recupera_a_otras_seis_familias-DLEC_67721
(accessed on 26 January 2024)

References

Ariel, Mira. 1990. *Accessing Noun Phrase Antecedents*. London: Routledge.

Ariel, Mira. 2001. Accessibility theory: An overview. In *Text Representation: Linguistic and Psycholinguistic Aspects*. Edited by Ted Sanders, Joost Schilperoord and Wilbert Spooren. Amsterdam: John Benjamins.

Bobaljik, Jonathan. 2012. *Universals in Comparative Morphology. Suppletion, Superlatives, and the Structure of Words*. Cambridge: MIT Press. [CrossRef]

Camacho, José. 1999. La coordinación. In *Gramática Descriptiva de la Lengua Española*. Edited by Ignacio Bosque and Violeta Demonte. Madrid: Espasa Calpe, vol. 1, pp. 2634–94.

Camacho, José. 2000. Structural restrictions on comitative coordination. *Linguistic Inquiry* 31: 366–75. [CrossRef]

Camacho, José. 2013. *Null Subjects*. Cambridge: Cambridge University Press.

Chomsky, Noam. 1981. *Lectures on Government and Binding*. Dordrecht: Foris.

Conti Jiménez, Carmen. 2005. Pluralidad comitativa. *Verba: Anuario Galego de Filoloxia* 32: 275–306.

Daniel, Michael, and Edith Moravcsik. 2013. The associative plural. In *The World Atlas of Language Structures Online*. Edited by Matthew Dryer and Martin Haspelmath. Leipzig: Max Planck Institute for Evolutionary Anthropology. Available online: http://wals.info/chapter/36 (accessed on 20 July 2023).

Eberenz, Rolf. 2000. *El Español en el Otoño de la Edad Media. Sobre el Artículo y los Pronombres*. Madrid: Gredos.

Embick, David, and Alec Marantz. 2008. Architecture and blocking. *Linguistic Inquiry* 39: 1–53. [CrossRef]
Feldman, Anna. 2002. On NP-Coordination. In *Yearbook 2002*. Edited by Sergio En Baauw, Mike Huiskes and Maaike Schoorlemmer. Utrecht: Utrecht Institute of Linguistics OTS, pp. 39–67.
Halle, Morris. 1997. Distributed Morphology: Impoverishment and Fission. *MIT Working Papers in Linguistics* 30: 425–49.
Holmberg, Anders, and Klaus Kurki. 2019. *We . . .with Anna*: Inclusory coordination in Finnish and Fenno-Swedish. In *The Sign of the V: Papers in Honour of Sten Vikner*. Edited by Ken Ramshøj Christensen, Henrik Jørgensen and Johanna L. Wood. Aarhus: Aarhus University, pp. 243–66. [CrossRef]
Ionin, Tania, and Ora Matushansky. 2002. DPs with a twist: A unified analysis of Russian comitatives. In *Proceedings of Formal Approaches to Slavic Linguistics 11: The Amherst Meeting*. Edited by Zeljko Boskovic, Steven Franks and William Snyder. Ann Arbor: Michigan Slavic Publications, pp. 255–74.
Kany, Charles. 1970. *Sintaxis Hispanoamericana*. Madrid: Gredos. First published 1945.
Kayne, Richard. 1994. *The Antisymmetry of Syntax*. Cambridge: The MIT Press.
Kratzer, Angelika. 1996. Severing the external argument from its verb. In *Phrase Structure and the Lexicon*. Edited by Johan Rooryck and Laurie Zaring. Dordrecht: Kluwer, pp. 109–37.
Kratzer, Angelika. 2009. Making a Pronoun: Fake Indexicals as Windows into the Properties of Pronouns. *Linguistic Inquiry* 40: 187–237. [CrossRef]
Lakoff, George, and Samuel Peters. 1969. Phrasal conjunction and symmetric predicates. In *Modern Studies in English: Readings in Transformational Grammar*. Edited by David Reibel and Sanford Shane. Englewood Cliffs: Prentice Hall, pp. 113–42.
Leonetti, Manuel. 2022. Topics and the interpretation of null subjects. In *Determiners and Quantifiers. Functions, Variation, and Change*. Edited by Chiara Gianollo, Klaus von Heusinger and Maria Napoli. Leiden: Brill, pp. 94–129.
Leonetti, Manuel, and Mª Victoria Escandell Vidal. 2020. Strong pronouns as postverbal subjects in Spanish and Italian. In *Theoretical Approaches to Contrastive Linguistics. Morphological and Syntactic Perspectives*. Edited by Michalis Georgiafentis, Giannoula Giannoulopoulou, Maria Koliopoulou and Angeliki Tsokoglou. Londres: Bloomsbury Academic.
López, Luis. 2009. *A Derivational Syntax for Information Structure*. Oxford: Oxford University Press.
Mare, María. 2012. *Sobre la Naturaleza de la Compañía*. Neuquén: EDUCO.
Mare, María. 2013. La concordancia comitativa en el español rioplatense. In *El español de la Argentina: Estudios Gramaticales*. Edited by Ángela Di Tullio. Buenos Aires: Eudeba, pp. 59–84.
Mare, María. 2015. *Proyecciones Funcionales en el Ámbito Nominal y Concordancia: Un Abordaje en Términos de Variación*. Munich: Lincom.
Mare, María. 2020. On plurality in verbs prefixed by COM. *Lingue Antiche e Moderne* 9: 157–84. [CrossRef]
Mare, María. 2023. A formal approach to Spanish "genitive" pronouns in non-nominal domains. *Languages* 8: 233. [CrossRef]
Mare, María, and Enrique Pato. 2017. La concordancia comitativa en español: Caracterización general y valores discursivos. *Revista Española de Lingüística* 47: 66–90.
Maslova, Elena. 2007. Reciprocal and polyadic: Remarkable reciprocals in Bantu. In *Reciprocal Constructions*. Edited by Vladimir Nedjalkov. Amsterdam and Philadelphia: John Benjamins, pp. 335–52.
Pascual Pou, Montserrat. 1999. *The 'Instrumental Phrase': Is It an Adjunct, an Argument or a Predicate?* Bellaterra: Universitat Autònoma de Barcelona, CatWPL.
RAE-ASALE. 2009. *Nueva Gramática de la Lengua Española*. 2 vols. Madrid: Espasa.
Rigau, Gemma. 1989. Prédication holistique et sujet nul. *Revue des langues Romanes* 93: 201–21.
Rigau, Gemma. 1990. The semantic nature of some Romance prepositions. In *Grammar in Progress*. Edited by Joan Mascaró and Maria Nespor. Dordrecht: Foris Publications, pp. 363–73.
Samek-Lodovici, Vieri. 1996. Constraints on Subjects: An Optimality-Theoretic Analysis. Ph.D. dissertation, Rutgers University, New Brunswick, NJ, USA.
Schwartz, Linda. 1988. Conditions on verb-coded coordination. In *Studies in Syntactic Typology*. Edited by Michael Hammond, Edith A. Moravcsik and Jessica Wirth. Ámsterdam: John Benjamins, pp. 53–73.
Sigurðsson, Halldor, and Jim Wood. 2020. "We Olaf": Pro[(x-)NP] constructions in Icelandic and beyond. *Glossa: A Journal of General Linguistics* 5: 16. [CrossRef]
Stassen, Leo. 2000. *And*-languages and *with*-languages. *Linguistic Typology* 4: 1–54. [CrossRef]
Stolz, Thomas, Cornelia Stroh, and Aina Urdze. 2006. *On Comitatives and Related Categories: A Typological Study with Special Focus on the Languages of Europe*. Berlín: Mouton de Gruyter.
Stolz, Thomas, Cornelia Stroh, and Aina Urdze. 2013. Comitatives and instrumentals. In *The World Atlas of Language Structures Online*. Edited by Matthew S. Dryer and Martin Haspelmath. Leipzig: Max Planck Institute for Evolutionary Anthropology. Available online: http://wals.info/chapter/52 (accessed on 20 July 2023).
Vassilieva, Masha. 2005. Associative and Pronominal Plurality. Ph.D. thesis, Stony Brook University, Stony Brook, NY, USA.
Vassilieva, Masha, and Richard K. Larson. 2005. The semantics of the plural pronoun construction. *Natural Language Semantics* 13: 101–24. [CrossRef]

languages

MDPI

Article

Code-Switching at the Interfaces

Antje Muntendam [1,*] and M. Carmen Parafita Couto [2,3,*]

[1] Department of Modern Languages and Linguistics, Florida State University, Tallahassee, FL 32306, USA
[2] Centre for Linguistics, Leiden University, 2300 RA Leiden, The Netherlands
[3] Language Variation and Textual Categorisation (LVTC), Universidade de Vigo, 36310 Vigo, Spain
* Correspondence: amuntendam@fsu.edu (A.M.); m.parafita.couto@hum.leidenuniv.nl (M.C.P.C.)

Abstract: One characteristic of multilingual speakers is that in everyday life, they may integrate elements from their languages in the same sentence or discourse, a practice known as code-switching. This paper examines code-switching at the interfaces, in particular as related to information structure. Despite the fact that a core question of modern linguistic theory is how syntactic and information-structural theories interact in accounting for licensing of different grammatical phenomena, there has been relatively little literature on code-switching and information structure. In this paper, we provide an overview of the available literature on code-switching across different language combinations, focusing in particular on subject pronoun–verb switches, ellipsis, light verbs, topic/focus particles, and code-switching between sign languages. We argue that the study of the interplay between information structure and code-switching sheds light on our understanding of multilingual grammars and language competence more generally. In this regard, we discuss theoretical and methodological considerations to guide future studies.

Keywords: code-switching; interfaces; information structure; light verbs; subject pronouns; intonation; ellipsis

Citation: Muntendam, Antje, and M. Carmen Parafita Couto. 2024. Code-Switching at the Interfaces. *Languages* 9: 258. https://doi.org/10.3390/languages9080258

Academic Editor: Ana Ojea

Received: 6 January 2024
Revised: 10 June 2024
Accepted: 13 July 2024
Published: 25 July 2024

1. Introduction

Multilingual speakers display an interesting characteristic in their everyday communication: the practice of code-switching, where elements from different languages seamlessly integrate into a single sentence or discourse (cf. Deuchar 2012). For example, in Spanish, the verb, *hacer*, meaning "do" or "make," can serve as both a lexical verb of creation and a causative verb. When integrated into Spanish–German code-switching, *hacer* is also employed as a light verb, losing a significant portion of its semantic content (1). However, this usage is confined to code-switching contexts (González-Vilbazo 2005; González-Vilbazo and López 2011).

(1)	*Vamos*	*a*	*hacer*	schreiben	*la*	Matharbeit.
	go.1PL	PREP	do.INF	write.INF	DET.FEM	maths.homework

We will write the maths homework.
(Adapted from González-Vilbazo 2005, p. 202)

In the following sections, we explore studies on code-switching at the interfaces, focusing specifically on its intersections with information structure (or information packaging), i.e., the way in which information is formally packaged within a sentence. The interface between grammar and discourse relates to how the organization of information within sentences (information structure) influences and is influenced by the syntactic form of sentences (syntax) and how these sentences fit into larger discourse contexts (discourse).[1] For an overview, see Erteschik-Shir (2007). The notion of information structure subsumes various dichotomies, such as topic/comment, focus/presupposition, theme/rheme, or background or given/new information, which divide a sentence into two parts based on pragmatic notions (Féry et al. 2007; Krifka 2008). Topic refers to what the sentence is about, whereas the rest of the sentence is the comment (Lambrecht 1996; Zubizarreta 1998). Focus

roughly refers to the new or non-presupposed information of a sentence, whereas the rest of the sentence is given, presupposed, or shared information (Chomsky 1971; Jackendoff 1972; Zubizarreta 1998). The literature presents conflicting views on the nature of focus, its distinction from topic, the types of foci, and their effects. For instance, a distinction is often made between broad and narrow focus. In broad focus, all information is new, and not one element is highlighted, whereas in narrow focus, one element of the sentence is highlighted. Furthermore, a distinction is often made between neutral focus and contrastive (or corrective) focus. In contrastive focus, one alternative is selected and contrasted with another alternative (cf. Erteschik-Shir 2007). Discrepancies also exist regarding the potential conflation of topic/comment with focus/presupposition (cf. Parafita Couto 2005 for an overview). Cross-linguistically, information structure can be expressed in syntax (e.g., word order), morphology (e.g., through topic and focus markers), and/or prosody (e.g., through intonation). Interestingly, in code-switching, we may find two languages that differ in their expression of information structure. This prompts an exploration of how such differences impact or limit the phenomenon of code-switching. In this paper, we acknowledge the diverse perspectives on information structure held by each author, striving to identify overarching patterns.

Within the realm of linguistic theory, an important question revolves around the dynamic interplay between syntactic and information-structural theories when explaining the mechanisms that underlie various grammatical phenomena (cf. Féry and Ishihara 2022). However, there exists a relative scarcity of literature that addresses the intricate relationship between code-switching and information structure. As such, in this paper, we undertake the task of providing an encompassing overview of the existing (albeit limited) body of research on code-switching across language combinations. Our focus is directed towards specific manifestations of code-switching, including light verbs (as discussed by González-Vilbazo and López 2011, 2012), subject pronoun–verb switches (as explored, for instance, by Bustin et al. 2024), ellipsis (examined by Merchant 2015; González-Vilbazo and Ramos 2019; Delbar 2022), and intonation (studied by Olson and Ortega-Llebaría 2010).

Code-switching has been studied from different perspectives, and different theoretical models have been proposed, with differing and sometimes contradicting predictions regarding the (un)acceptability of code-switches. In Section 2, we discuss the main theoretical approaches to code-switching. In Section 3, we provide an overview of different studies, contextualizing the findings associated with information structure and code-switching, which stem from diverse theoretical frameworks (e.g., generativism and the matrix language framework). We acknowledge the limitations of the existing research, which often feature examples of code-switching studied in isolation and out of context, i.e., some do not explicitly consider information structure. Moreover, many of these studies tend to involve a relatively small number of participants. In light of these limitations, Section 4, following this overview, presents a discussion of the theoretical and methodological considerations that should guide future studies in this area.

Overall, this contribution endeavors to shed light on the scope of the existing literature concerning information structure and code-switching at the interfaces of syntax, prosody, and discourse. As such, we aim to provide a holistic understanding of code-switching at the interfaces (syntax–prosody–discourse), examining the available sources to delineate the existing landscape and identify key themes, gaps, and potential avenues for further research.

2. Theoretical Approaches to Code-Switching

In the study of code-switching, the focus is on understanding the boundaries that allow or disallow the mixing of languages in speech. The study of code-switching is relatively recent, originating in the early 1970s, opposing earlier assertions of randomness in language switching (cf. Weinreich 1953; Labov 1971). This early research laid the groundwork by providing a descriptive foundation that outlined constraints on language mixing. For instance, Gumperz and Hernández-Chávez (1970) noted "linguistic constraints" on

pronominal pronouns in Spanish–English code-switching, while Aguirre (1976) investigated grammatical judgments on code-switching, marking the overt emergence of theories of code-switching. As such, linguists at this time, such as Lipski (1978), Pfaff (1979), and Poplack (1980), argued that code-switching between languages is not arbitrary but is governed by structural constraints. In these early years, researchers concentrated on identifying the specific structural constraints characterizing this bilingual speech phenomenon.

Pfaff (1979) conducted a pioneering study on code-switching, emphasizing that the mixture of Spanish and English is socially motivated but constrained by clear linguistic principles. Based on an analysis of conversational data from around 200 speakers, Pfaff argued against the need for a third grammar and proposed that Spanish and English grammars are intertwined, based on various constraints. She identified structural constraints favoring surface structures common to both languages. For instance, Spanish auxiliaries and English participles can be combined, while the mixing of adjectives and nouns within noun phrases is restricted due to differing surface orders in Spanish and English. This observation aligns with Poplack's (1980) "Equivalence Constraint," which suggests that code-switching tends to occur where the juxtaposition of elements from both languages does not violate syntactic rules. It is important to highlight that the constraints postulated by Poplack focused on surface switch points rather than on the material being switched. More recent studies suggest that exclusively considering switch points is not adequately explanatory and that the morphosyntax of the entire clause needs to be taken into account (cf. Deuchar 2020; Parafita Couto et al. 2023).

Given instances that contradict Poplack's constraints, Poplack (2001) proposed replacing the idea of constraints on code-switching with that of general principles. One such principle involves an asymmetry between the contributions of the two languages in code-switching. This was already observed by Joshi (1982), who presented evidence of such asymmetry based on Marathi–English data. He distinguished between the bilingual's "matrix" and "embedded" languages and argued that the matrix language serves as the source for both open and closed class items (e.g., determiners, prepositions, tense-marked verbs) in bilingual speech, whereas the embedded language can only contribute open class items (e.g., nouns, adjectives). This differentiation between the matrix and embedded language in code-switching served as the foundation for Myers-Scotton's (1993, 2002) Matrix Language Frame (MLF) model. Myers-Scotton encapsulates the mentioned asymmetry in her "Asymmetry Principle," characterizing it as the "morphosyntactic dominance of one variety in the frame" of a clause within bilingual speech (Myers-Scotton 2002, p. 9). The language providing the morphosyntactic framework is termed the "matrix language," while the other language becomes the "embedded language," with its material constituting switches away from the matrix language. While there are no restrictions on the grammatical categories of constituents from the matrix language, only certain broadly defined "open class" items can be borrowed from the embedded language. An exception to this limitation occurs when a "chunk" of embedded language items appears together, as in embedded language islands.

The 4-M model, introduced by Myers-Scotton and Jake (2000), builds upon the MLF model and expands its scope. It attempts to explain the variation in the frequency of different morpheme types in code-switched data. The model classifies morphemes into four types: content morphemes, early system morphemes, late system morphemes, and outsider system morphemes. Content morphemes participate in the thematic structure of a clause by receiving or assigning thematic roles. Examples include nouns and pronouns occurring in argument positions. Early system morphemes, such as determiners and affixes modifying their heads, are characterized by their involvement in the lexical–conceptual level. They contribute semantic and pragmatic features like definiteness, plurality, and aspect. Early system morphemes are conceptually activated and depend on their directly elected content morpheme heads for form and meaning information. Late system morphemes, according to the System Morpheme Principle, are supplied by the matrix language. They include elements like agreement morphology and case affixes that make relationships within the

clause more transparent, especially between arguments and predicates. Outsider system morphemes do not interact with structure-building processes at the functional or positional levels unless they are involved in case assignment. We will return to the distinction between content morphemes and system morphemes in our discussion of information structure and subject pronoun–verb switches.

Constraint-based approaches, including the Matrix Language Framework model, faced challenges from the perspective advocating a Null Theory, suggesting that no specific restrictions should apply exclusively to code-switching. Pfaff (1979, p. 314) had already argued that there is no need to postulate the existence of a third grammar to explain mixed-language utterances. Di Sciullio et al. (1986, p. 7) characterized code-switching as an ordinary aspect of language use, requiring no special stipulation, while Joshi (1982, p. 150) asserted that there is no third grammar in code-switched speech. In contrast, López (2020) distinguishes between two types of Null Theories: Broad Null Theory, utilizing Universal Grammar constraints to explain apparent code-switching constraints, and Narrow Null Theory, exclusively relying on grammatical features of the participating languages.[2]

These challenges to code-switching-specific constraints were often voiced within the generative linguistic framework, which views language structure as multi-layered and hierarchical. The initial explicit Chomskyan approaches to code-switching emerged during the Government and Binding era (Chomsky 1981a, 1981b). Examples include studies by Joshi (1982), Bentahila and Davies (1983), Woolford (1983), and Di Sciullio et al. (1986). These approaches signaled a departure from Poplack's linear perspective, adopting a more hierarchical approach. For instance, MacSwan (1999, 2005, 2009) proposes a code-switching model within Minimalism where lexical items are inserted at the beginning of structure building, suggesting a universal structure-building process for all speakers. Lexical items are chosen from each contributing language's lexicon to compose the Numeration. The construction of structures involves merging elements found in the Numeration. During the derivation, lexical items merge according to the features that require checking and valuation before transferring to the interfaces upon completing a phase. Notably, the bilingual structure's computational process aligns with that of monolingual structures. As such, MacSwan (2005, p. 75) illustrates the I-language of a bilingual speaker by feeding a computational system (CHL) with lexical items from different languages (Lx and Ly). For a thorough exploration of the differentiation between "constraint-based" and "constraint-free" theories of code-switching, readers are directed to MacSwan (2014), who introduced this distinction.

Some studies have attempted to evaluate the predictions of different theoretical models, in particular the theoretical predictions of the MLF model and the Minimalist Program (Parafita Couto and Gullberg 2019; Herring et al. 2010; Parafita Couto et al. 2015; Vaughan-Evans et al. 2020, among many others). The results of these studies mostly align either with both accounts or remain inconclusive, making it challenging to distinguish between the predictions of both theories in naturalistic data. In contrast, several of the studies discussed in this paper contrast the two theories, using (semi-)experimental data.

Indeed, the tension may arise from the fact that linguistic communities and individual speakers differ in their language use. By recognizing and describing this variability, researchers can paint a more comprehensive picture of multilingual language practices. Understanding the spectrum of choices among communities and speakers adds necessary depth to theoretical development, offering insights that go beyond the confines of conflicting theoretical perspectives. Muysken (2000, 2013) stands out as a pioneer in recognizing the impact of various factors in the development of his code-switching typology. He introduces four types of language mixing in his code-switching typology, each reflecting varying degrees of contribution from lexical items and structures of two languages. The specific outcome is influenced by numerous factors, including typological distance, political distance, and community norms. Similarly, Bhatt and Bolonyai (2011) incorporate socio-cognitive constraints into their Optimality Theory (OT) framework of bilingual grammar. Furthermore, Goldrick et al. (2016) propose Gradient Symbolic Computation (GSC) as a formalism

to explain the systematic nature of code-switching patterns, integrating psycholinguistic concepts of bilingual co-activation with generativist perspectives on grammar. Finally, a more recent perspective from Aboh and Parafita Couto (2024) advocates for a paradigm shift that recognizes the intricate interplay of linguistic features, hybridity, community norms, and multilingualism to foster a more holistic understanding of language. They propose studying diverse ecologies[3] where multilingual acquisition occurs, integrating factors like experience and cognition into the development of models of human language capacity. We come back to this in the discussion.

This concise theoretical overview lays the groundwork for the forthcoming section. Given the brevity required here, for a recent overview of theoretical approaches to code-switching, please refer to Parafita Couto et al. (2023). In the ensuing discussion, we will provide an overview of diverse studies that have examined the intersection of code-switching and information structure within a particular theoretical framework. These studies investigate the complex relationship between code-switching phenomena and the organizational principles governing the information structure within mixed utterances. By exploring the findings and insights derived from these diverse theoretical lenses, we aim to contribute to a deeper understanding of the complex interplay between code-switching and information structure in linguistic research.

3. Studies on Code-Switching at the Interfaces

Interface phenomena involve the interaction of different linguistic domains (e.g., syntax, morphology, phonology, pragmatics). Studies on code-switching at the interfaces are informative in two ways. By taking into account strategies to encode information structure of the participating languages, researchers can test hypotheses that cannot be tested with monolingual data alone. Moreover, considering the information structure of the sentences gives us a more complete picture of when code-switches are acceptable. This section includes studies on light verbs (González-Vilbazo and López 2011, 2012), subject pronoun–verb switches (Bustin et al. 2024), intonation (Olson and Ortega-Llebaría 2010), ellipsis (Delbar 2022; González-Vilbazo and Ramos 2019; Merchant 2015), topic and focus particles, and code-switching between sign languages (Stoianov et al. 2023).

3.1. Light Verbs

An interesting phenomenon of code-switching at the interfaces concerns light verbs, exemplified by the following Spanish–German example from González-Vilbazo and López (2012), which uses a generativist (Minimalist Program) framework:[4]

(2) *Hizo* nähen das Hemd. (González-Vilbazo and López 2012, p. 35)
did sew the shirt
(He/she) sewed the shirt.

In (2), a form of the Spanish light verb, *hacer*, "do," is used with the German lexical verb, *nähen*, "to sew," in the infinitive. This type of code-switch has been attested in numerous other language pairs. In fact, González-Vilbazo and López (2012) suggest that it occurs in most if not all code-switching pairs. The paper proposes the analysis in Figure 1 for this construction (González-Vilbazo and López 2012, p. 35), where the light verb in little *v* takes a VP as its complement, which has a lexical V as its head.[5]

In this construction, the little *v* and the lexical verb come from two different lexica, which allows for the study of the properties of little *v*. González-Vilbazo and López (2011, 2012) argue that the light verb can only come from one of the languages, which in Spanish–German code-switching is Spanish.[6] The XP, however, can come from either language.

Following Chomsky's phase theory, González-Vilbazo and López argue that *v*P is a phase and that little *v* is the head of the phase.[7] As the head of the phase, little *v* determines some grammatical properties of the VP. In particular, González-Vilbazo and López propose that little *v* determines the word order (XP V/V XP) of the VP, its prosodic phrasing, and the encoding of information structure (in particular, focus/background). Code-switching between two typologically different languages lends itself well to testing this hypothesis.

Based on conversational data and oral and written grammaticality judgment tasks from over 85 Spanish–German bilinguals living in Spain, González-Vilbazo and López (2011, 2012) showed that in constructions with a Spanish light verb and a German lexical verb, word order, prosodic phrasing, and the encoding of focus/background follow Spanish.

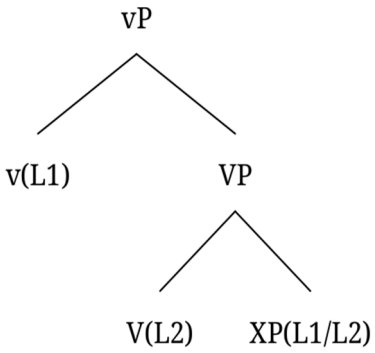


Figure 1. Analysis of the light verb construction (González-Vilbazo and López 2012, p. 35).

First, regarding the word order of the VP, in German, the word order of constructions with an auxiliary verb is OV, whereas in Spanish, it is VO. As shown in (3), the light verb constructions in Spanish–German code-switching follow Spanish VO word order rather than German OV word order. That is, the word order is governed by the Spanish light verb in little *v*.

(3) *Juan* *ha* *hecho* verkaufen die Bücher. (González-Vilbazo and López 2012, p. 42)
Juan has done sell the books
Juan has sold the books.

Second, regarding the prosodic structure of the VP, Spanish and German differ in their prosodic phrasing. In German, the lexical verb and its complement can appear in one prosodic phrase (with one pitch accent, on the object), whereas in Spanish, they appear in two separate prosodic phrases (with two pitch accents). In the light verb construction in Spanish–German code-switching, the lexical verb and its complement are in two separate prosodic phrases, as in Spanish (4a). Example (4b), where the lexical verb and its complement appear in one prosodic phrase with one accent, is ungrammatical.

(4) a. *Juan* *ha* *hecho* (ϕ verKAUfen) (ϕ die BÜcher).
Juan has done sell the books
Juan has sold the books.
b. #*Juan* *ha* *hecho* (ϕ verkaufen die BÜcher).
Juan has done sell the books
Juan has sold the books. (González-Vilbazo and López 2012, p. 43)

Of particular relevance to the current paper, little *v* determines the encoding of background information, i.e., whether it is deaccented or dislocated. In Spanish, background information (i.e., given information, or topic) is typically moved to the left periphery of the sentence and doubled by a clitic.[8] In German, background information can be topicalized, where the object is moved to SpecC, or it can be deaccented. In Spanish–German code-switching, the background information is dislocated as in Spanish, which is exemplified in (5a), in which the object, *die Uhren*, "the watches," is moved to the left periphery of the sentence and doubled using a Spanish clitic with the same φ-features. Specifically, the Spanish clitic is feminine and plural, like *die Uhren*, "the watches" in German. Example (5b), in which the object is deaccented and appears in a single prosodic phrase with the lexical verb (as in German), is not acceptable in the context provided. This sentence would only be possible if the verb were in contrastive focus. Moreover, example (5c), in which the object is moved to the beginning of the sentence, but which lacks a clitic (as in German), is not possible in this context, in which *die Uhren*, "the watches" is background information. This sentence would also only be possible in a context in which the object is in contrastive focus.

(5) [Context: What happened to the watches?]

a. *Juan* die Uhren *las* *hizo* verKAUfen.
Juan the watches CL.ACC did sell
The watches, Juan sold them.

b. #*Juan* *hizo* verKAUfen die Uhren.
Juan did sell the watches
Juan sold the watches

c. #Die Uhren *hizo* *Juan* verkaufen.
the watches did Juan sell
The watches, Juan sold them. (González-Vilbazo and López 2012, p. 44)

In sum, these examples show that the encoding of information structure (focus/background) in the light verb construction in Spanish–German code-switching is as in Spanish. This can be explained if it is assumed that little *v* determines the properties of the VP. An important contribution of this study is that it shows that little *v* determines more properties than what has been argued in the literature so far. Crucially, this could not have been shown based on monolingual data alone.

González-Vilbazo and López (2011) argue that the Matrix Language Frame model cannot account for the light verb construction. According to the Matrix Language Frame model, the language of the inflection of the verb is the matrix language and determines the word order of the sentence. For instance, in (6), the light verb *hizo* determines that the matrix language of the sentence is Spanish, and thus the word order within the VP is VO.

(6) *Juan* *hizo* nähen das Hemd. (González-Vilbazo and López 2012, p. 35)
Juan did sew the shirt
Juan sewed the shirt.

However, González-Vilbazo and López (2011) show that the word order is only VO when there is a light verb. Sentences with code-switches between a Spanish lexical verb and a German complement clause (7) or a DP (8) follow German word order:[9]

(7) a. *Juan* *dijo* dass Johannes klug ist.
Juan said that Johannes clever is
Juan said that Johannes is clever.

b. **Juan* *dijo* dass Johannes ist klug. (González-Vilbazo and López 2011, p. 847)
Juan said that Johannes is clever
Juan said that Johannes is clever.

(8) a. *Juan* *vio* die kluge Frau.
Juan saw the clever woman
Juan saw the clever woman.

b. **Juan* *vio* die Frau kluge. (González-Vilbazo and López 2011, p. 847)
Juan saw the woman clever
Juan saw the clever woman.

In (7), the lexical verb is in Spanish and the complement clause is in German. The sentence is only acceptable if the verb in the complement clause is in final position (7a), as in German. Sentence (7b), in which the verb is in a non-final position, is unacceptable. In (8), a Spanish lexical verb is followed by a German DP. The contrast between (8a) and (8b) shows that the only acceptable word order is Adjective–Noun, as in German. Importantly, the Matrix Language Frame model cannot explain the difference in word order after lexical verbs versus light verbs. González-Vilbazo and López, on the other hand, have a straightforward and elegant explanation for these examples based on phase theory. Because the lexical verb and its complement belong to different phases, the lexical verb does not determine the properties of the complement, unlike light verbs.

More recent work on light verbs, however, has shown that the use of *hacer* as a light verb is not always available. Specifically, Balam et al. (2020) showed that light verbs with *hacer* are produced and accepted in Belize and New Mexico, but not in Puerto Rico, showing a variation across code-switching communities. This means that the production and acceptability of *hacer* + Vinf are shaped by community-specific code-switching practices, in the same way that differences are observed in monolingual language contexts. As

pointed out by an anonymous reviewer, these practices lead to the evolution of structural choices over time. Regarding the specific case of Belize, Balam et al. (2021) propose that a previously established Spanish-Maya template (such as *hacer lit'í*, "to stand on one's toes") in earlier generations, probably played a role in the spread and standardization of this hybrid structure as the community shifted from Spanish–Maya to Spanish–English bilingualism. Such findings emphasize the need to study code-switching from a language ecological perspective (cf. Mufwene 2014), recognizing the influence of community-specific practices on bilingual grammars.

3.2. Subject Pronoun–Verb Switches

Code-switching between subject pronouns and finite verbs (9a) has been claimed to be relatively infrequent in corpora and dispreferred in judgments (cf. Lipski 1978, 2019; Timm 1975; van Gelderen and MacSwan 2008). In contrast, switches between a full lexical DP and a finite verb are considered acceptable, as seen in (9b):

(9) a. **She* odia los exámenes. (Fernández-Fuertes et al. 2016, p. 80)
she hates the exams
She hates exams.

b. *That teacher* odia los exámenes. (Fernández-Fuertes et al. 2016, p. 80)
that teacher hates the exams
That teacher hates exams.

This has been confirmed in acceptability judgment tasks (Fernández-Fuertes et al. 2016; Lipski 2017). However, Lipski (2017) found that the accuracy for switches of lexical nouns and subject pronouns was similar in a concurrent memory-loaded repetition task, suggesting their acceptability. Several studies on code-switches between a subject pronoun and a verb have shown an effect of grammatical person (Bellamy et al. 2022; Fernández-Fuertes et al. 2016). Specifically, switches of third person pronouns are typically rated higher in an acceptability judgment task than switches of first and second person pronouns. Importantly, however, these previous studies on switches between a subject pronoun and a finite verb do not consider the information structure of the sentences. Sentences containing these switches are either presented in isolation, or the subject pronouns have referents that were previously mentioned. For instance, in Fernández-Fuertes et al. (2016), the lexical NPs and the subject pronouns were always given information. Specifically, they appeared as an answer to a *wh*-question, as in *¿Qué hace el chico?*, "what does the boy do?"—*The boy bebe agua*, "The boy drinks water." However, the information structural status of subject pronouns may play a role in the acceptability of these switches. For instance, González-Vilbazo and Koronkiewicz (2016) found a difference between prosodically stressed subject pronouns in contrastive focus (as indicated by capital letters in their acceptability judgment task) and other subject pronouns. Specifically, switches of stressed subject pronouns (in contrastive focus) were rated as highly acceptable, whereas unmarked subject pronouns were not, suggesting an effect of information structure.

Information structure, in particular, focus, is highly relevant because it allows us to distinguish between the predictions of different theoretical models of code-switching, as shown in Bustin et al. (2024). This study focused on Spanish–English code-switches between a pronominal subject and a finite verb, as in (10), and used the notion of focus to test two code-switching models: a Minimalist Approach to code-switching (MacSwan 1999, 2000; van Gelderen and MacSwan 2008) and the MLF/4-M model (Myers-Scotton 1993; Myers-Scotton and Jake 2000).

(10) Diego listens to music while he *corta el césped.*
Diego listens to music while he cuts the grass.
(Bustin et al. 2024, suppl. materials)

Focus plays an important role in pronoun use in Spanish, which is a null subject language (unlike English). In broad focus, where the entire sentence is new information, and no single element is highlighted, null pronouns are used (11a). However, in contrastive focus, where one element is highlighted and contrasted with alternatives, overt pronouns are used

(11b) (for more on the use of null and overt pronouns and its relation to information structure, see, for instance, Jiménez-Fernández 2016; Frascarelli and Jiménez-Fernández 2019):

(11) a. Diego$_{\mathbf{i}}$ escucha música mientras pro$_{\mathrm{i}}$ corta el césped.
Diego listens to music while he cuts the grass.

b. Diego$_{\mathbf{i}}$ escucha música mientras ÉL$_{*\mathrm{i/j}}$ corta el césped.
Diego listens to music while he cuts the grass.
(Bustin et al. 2024, suppl. materials)

This distinction is relevant for the predictions of the MLF/4-M model. The MLF/4-M model allows for code-switching when the pronouns in both languages (here Spanish and English) are obligatory. According to this model, null pronouns in Spanish are system morphemes, whereas overt pronouns in contrastive focus are content morphemes. English always requires overt pronouns, which are classified as content morphemes. According to the MLF/4-M model, then, in contrastive focus sentences, code-switching between a subject pronoun and a finite verb is permitted, as the pronouns in both languages are content morphemes and obligatory. In broad focus sentences, this type of code-switching is not allowed as the (null) pronoun in Spanish is a system morpheme. The Minimalist Approach to code-switching as proposed by MacSwan (1999, 2000) and van Gelderen and MacSwan (2008) does not distinguish between pronouns in broad focus and contrastive focus sentences. According to this approach, code-switching between a subject pronoun and a finite verb is generally predicted to be unacceptable. The predictions of the two models are summarized in Table 1 adapted from Bustin et al. (2024, p. 316).

Table 1. Predictions for Spanish–English code-switching between subject pronouns and finite verbs (* = ungrammatical and $\surd$ = grammatical).

Sentence Type	Pronoun Type for Each Language	Minimalist Program Prediction (van Gelderen and MacSwan 2008)	MLF/4-M Prediction
Broad focus	Spanish null pronoun English overt pronoun	*	*
Contrastive focus	Spanish overt pronoun English overt pronoun	*	$\surd$

It should be noted that the Minimalist Program could predict code-switching between a subject pronoun and finite verb to be acceptable if overt subject pronouns in contrastive focus would be considered DPs. van Gelderen and MacSwan (2008) argue that subject pronouns are base-generated under D and then move to a D position in the Spec of T. Because this creates a mixed language head, violating the PF Interface Condition, code-switching between subject pronouns and finite verbs is unacceptable. However, Bustin et al. (2024) propose that if subject pronouns in contrastive focus were considered DPs, they would move to the Spec of TP, like lexical subjects, and would be acceptable. Alternatively, it could be argued that subject pronouns in contrastive focus move to the left periphery of the sentence and appear as a lexical head.

The contrasting predictions of the two models, as presented in Table 1, allowed Bustin et al. (2024) to experimentally test the two models, using a concurrent memory-loaded repetition task, adapted from Lipski (2017).[10] In this task, the participants first heard a sequence of four numbers, then saw a picture, and then heard a sentence, followed by a beep. After the beep, the participants repeated back the sequence of numbers and the sentence. Crucially, the specific picture provided a context for the sentence. Specifically, a picture in which one person performed two actions was used to elicit a contrastive focus interpretation, whereas a picture, in which two people performed the two actions, elicited a broad focus interpretation. Accuracy in repeating back the sentence was measured. In this study, accuracy was considered to reflect acceptability, whereas modifications to the structure were interpreted as that structure not being acceptable.

The participants completed the task in both unilingual Spanish and code-switching mode. As there is variation in pronoun use among Spanish varieties and among Spanish–

English bilinguals, unilingual mode was used to ensure that the participants had the expected pronoun use in Spanish, that is, null pronouns in broad focus and overt pronouns in contrastive focus. The participants for the code-switching analysis were 38 adult Spanish–English early bilinguals, who showed the expected pronoun use in Spanish.

The results revealed a higher accuracy for sentences with a Spanish overt subject pronoun and an English finite verb in contrastive focus than in broad focus. These results provide support for considering the information structure of the utterances. Moreover, accuracy was higher in the contrastive focus condition than in the broad focus condition, against the Minimalist Program as proposed by van Gelderen and MacSwan (2008), which predicted code-switching to be unacceptable in both contexts. Finally, accuracy was higher for the subject pronoun–finite verb switches than for the distractors in this study, which included switches after a preposition, complementizer, conjunction, or auxiliary verb. This provides additional support for the permissibility of these code-switches.

All in all, the results showed that code-switching between a subject pronoun and finite verb is more complicated than previously thought. The study provides empirical support for a distinction between overt and null pronouns and shows the importance of considering the information structure of sentences. Specifically, the manipulation of information structure provided a better understanding of the permissibility of code-switching between a subject pronoun and a verb. Moreover, by looking at Spanish unilingual pronoun use, Bustin et al. were able to carefully select participants and test the predictions of the two models.

In the code-switching literature, the information structure of sentences is usually not taken into account. However, it might explain the low frequency of code-switches between a subject pronoun and a finite verb in corpora and the low acceptability in studies that do not manipulate information structure and often present sentences in isolation. Additionally, other studies do not control for Spanish unilingual pronoun use, possibly obscuring the results. All in all, when we only look at the syntax and not interfaces, we are missing an important fact of the permissibility of code-switching.

3.3. Intonation

In unilingual mode, intonation can be used in languages such as English and Spanish to convey the information structure of an utterance. However, relatively little is known about the use of intonation to mark information structure in bilingual, code-switching mode. One of the few studies on the intonation of code-switching is Olson and Ortega-Llebaría (2010), who studied the effect of intonation on the perception of narrow contrastive focus (where one element in the utterance is highlighted and contrasted with previous information) in code-switched utterances. Their hypothesis was that code-switched utterances would be interpreted as narrow contrastive focus utterances more frequently than non-code-switched utterances, especially when there were no intonational cues.

Their perception experiment included three groups of participants with six participants per group: late bilinguals with English as their L1, late bilinguals with Spanish as their L1, and early (simultaneous) bilinguals. The latter group was exposed to or used code-switching. The experimental stimuli consisted of unilingual Spanish and code-switched utterances in broad focus (i.e., the entire utterance consists of new information), as in (12a) and (12b), respectively:

(12) a. Yo miro la luna de María. (Olson and Ortega-Llebaría 2010, p. 61)
I see María's moon.

b. Yo miro *the mommy Of* *María.* (Olson and Ortega-Llebaría 2010, p. 61)
I see María's mommy.

These stimuli were manipulated for pitch range and peak alignment, as narrow contrastive focus in Spanish is associated with a larger pitch range and earlier peak alignment than broad focus. Pitch range here refers to the difference between the peak that is associated with the stressed syllable and the preceding valley, and peak alignment refers to the location of the peak with respect to the stressed syllable.

The participants listened to an utterance and had to indicate whether the utterance was a response to the question in (13a), which elicits a broad focus utterance, or to the question in (13b), which elicits a narrow contrastive focus utterance. The participants were instructed to press a key if the utterance was a response to the question in (13b).

(13) a. ¿Qué pasa? (eliciting a broad focus utterance)
What happens?

b. ¿Miras el padre de María? (eliciting a narrow contrastive focus utterance)
Do you see Maria's father?

The results revealed that code-switched utterances were interpreted as narrow contrastive focus utterances more frequently than non-code-switched utterances, by all the groups. Importantly, this difference was larger (and significant only) when there were no intonation cues for narrow contrastive focus such as a larger pitch range. This means that code-switching by itself was perceived as a cue for narrow contrastive focus.

This study is one of the few studies looking at the interaction of prosody, focus, and code-switching, and it provides support for code-switching as a way to highlight or emphasize part of an utterance. However, it is based on a highly controlled experiment with manipulations of two utterances, which raises the question of ecological validity.[11] Further research is needed on the interaction of prosody, focus, and code-switching in both experimental and naturalistic settings.

In this section, we saw that code-switching and intonation are used to highlight information. In the next section, we discuss what happens when elements or sounds are omitted.

3.4. Ellipsis

Ellipsis involves intricate connections with information-structural concepts, as highlighted by Winkler (2022). Generally, ellipsis pertains to the deliberate omission of linguistic elements, encompassing both structure and sound. This omission is intrinsically linked to the idea of givenness, i.e., the unspoken or deleted segment is previously mentioned or understood. The remnants of the ellipsis site, which occur to the left or right of the omitted material, are frequently connected to the notion of contrastive topic and focus. In contemporary linguistic theory, a fundamental question revolves around the synergy between syntactic and information-structural theories in explaining the licensing mechanisms governing different types of ellipsis (Winkler 2022). Notably, discourse factors and information structure exert a profound influence on the form and interpretation of ellipsis, highlighting the essential role of information structure. However, although research on code-switching as well as ellipsis has witnessed remarkable growth over the past two decades, the specific realm of ellipsis within codeswitching remains an area that has not received commensurate attention (González-Vilbazo and Ramos 2019).

The study of ellipsis in code-switching is interesting because it allows us to test theories of ellipsis, and it sheds light on the identity relationship between the ellipsis and its antecedent. There are different theories of ellipsis, including the copy theory and deletion theory. Most studies on code-switching provide support for the deletion theory, according to which the syntactic structure of the ellipsis is deleted at Spell-Out or PF. Regarding the identity relation between the antecedent and the ellipsis, the question is whether it is semantic, syntactic, or a hybrid. Most studies on code-switching and ellipsis suggest that the relationship has to be hybrid, i.e., both semantic and morphosyntactic.

One recent study showing a hybrid relationship is González-Vilbazo and Ramos (2019), who report on a study on sluicing and code-switching with six participants. In sluicing or TP ellipsis, the entire *wh*-clause except for the *wh*-phrase is deleted. González-Vilbazo and Ramos tested connectivity effects in Spanish–German code-switching, in particular regarding case. In their study, they used verbs that assign different case in Spanish and German. For instance, the verb, *amenazar*, "to threaten," assigns accusative case in Spanish, but the German equivalent, *drohen*, assigns dative case. In code-switched sentences, the *wh*-phrase has the case assigned by the verb in the subordinate clause, that is, dative in (14) and accusative in (15):

(14)

Juan	*amenazó*	*a alguien*	aber	ich	weiss
Juan	threatened	someone.ACC	but	I	know
nicht	{*wen/	wem}	er	gedroht	hat.
not	who.ACC	who.DAT	he	threatened	has

Juan threatened someone, but I don't know who he threatened.

(González-Vilbazo and Ramos 2019, p. 11)

(15)

Juan	*amenazó*	*a alguien*	aber	ich	weiss
Juan	threatened	someone.ACC	but	I	know
nicht	{wen/	*wem}	*Juan*	*amenazó.*	
not	who.ACC	who.DAT	Juan	threatened	

Juan threatened someone, but I don't know who he threatened.

(González-Vilbazo and Ramos 2019, p. 11)

In sluicing, however, the *wh*-phrase has to have accusative case, as shown in (16):

(16)

Juan	*amenazó*	*a*	*alguien*	aber	ich	weiss	nicht
Juan	threatened	someone.ACC		but	I	know	not
{wen/	*wem}.						
who.ACC	who.DAT						

Juan threatened someone, but I don't know who.

(González-Vilbazo and Ramos 2019, p. 12)

This example shows that the elided verb was the Spanish verb, *amenazar*, and sheds light on the identity relation between the ellipsis and the antecedent. Specifically, it shows that the identity relation is not only semantic but also morphosyntactic. If the relation were only semantic, the use of *wem* in (16) should be acceptable, as the ellipsis and antecedent are semantically identical. However, the case assigner of the *wh*-phrase needs to be identical, which supports a hybrid analysis of ellipsis.

González-Vilbazo and Ramos (2019) cite further support from Nee (2012) on sluicing in Spanish–Zapotec code-switching. Some verbs in Spanish and Zapotec assign different cases. For instance, in Spanish, *hablar*, "to talk," selects a PP (*habló con*), whereas the Zapotec equivalent, *gunien*, is a transitive verb. Code-switched data show that (17) is grammatical because the Spanish verb, *habló*, selects for a PP, which in this case is a Zapotec PP (*tu cun*). Example (18) is unacceptable, because it involves P-stranding, which is not permitted in Spanish (nor Zapotec). Finally, the sentence in (19) is acceptable because *gunien* is a transitive verb and selects a DP. Nee concludes that the identity relation between the antecedent and the ellipsis is not only semantic but also includes lexical identity.

(17)

Juan	*habló*	*con*	*alguien*	per	kednanadia	tu	cun_i	<*habló* t_i>
Juan	spoke	with	someone	but	not.know.1S	who	with	spoke

Juan spoke with someone, but I don't know who.

(González-Vilbazo and Ramos 2019, p. 14)

(18)

**Juan*	*habló*	*con*	*alguien*	per	kednanadia	tu	<*habló con*>
Juan	spoke	with	someone	but	not.know.1S	who	spoke with

Juan spoke with someone, but I don't know who.

(González-Vilbazo and Ramos 2019, p. 14)

(19)

Juany	gunien	*pero*	*no*	*sé*	*quién*	<gunien>
Juan	spoke	but	not	know.1SG	who	Spoke

Juan spoke, but I don't know who to.

(González-Vilbazo and Ramos 2019, p. 14)

Another code-switching study that shows that the identity relation between the antecedent and the ellipsis is not purely semantic is Merchant (2015) on VP ellipsis in Greek–English code-switching. Merchant provides the example in (20), which contains a question–answer pair:

(20) a.

Píres	tin	tsánda	mazí	su?	(Merchant 2015, p. 204)
took.2S	the	bag	with	you	

Did you take the bag with you?

b. Yes, I did.

The response in (20b) is an example of ellipsis in code-switching, but interestingly, the data in (21) show that parallel responses without ellipsis (and the intended meaning "yes, I did take the bag with me") are not possible:

(21)	a.	*Yes,	*I*	*did*	píra	tin	tsánda	mazí	mu.
		yes	I	did	take.ACT.PERF.PAST.1S	the	bag	with	me
								(Merchant 2015, p. 204)	
	b.	*Yes,	*I*	*did*	pern	tin	tsánda	mazí	mu.
		yes	I	did	take[stem.form]	the	bag	with	me

Example (21a) is unacceptable, because the English auxiliary, *did*, cannot appear with a Greek inflected form, here *píra* "take.ACT.PERF.PAST.1S." The inflected form can only be created when a Greek root appears with a T node. On the other hand, (21b) is unacceptable because *pern*, "take," is a bare stem form of the verb, and bare stem forms cannot appear as free-standing words in Greek, which Merchant (2015) proposes is a morphological issue. Merchant (2015) argues that these examples show that the identity relation between the antecedent and the ellipsis needs to be morphological as well. To explain these and other examples, he proposes morphological elliptical repair. Together, all these code-switching studies show that the identity relation between the ellipsis and the antecedent needs to include morphosyntax, in addition to semantics.

All in all, most of the studies on code-switching in ellipsis show support for the deletion theory and suggest a hybrid relation between the ellipsis and the antecedent. However, Delbar (2022) did not find evidence for a hybrid relation in a study on NP ellipsis and gender agreement in Belgian Dutch–French code-switching. This study explored whether the choice of the grammatical gender showed a morphosyntactic link between the French elided noun and the Belgian Dutch antecedent, based on data from 23 Belgian Dutch–French bilinguals who participated in a two-alternative forced choice task. The materials for the task included sentences such as those in (22):

(22)	a.	Ik	eet	den	roden	appel	*et*	*tu*
		I	eat	the.M	red.M	apple.M	and	you
		manges	*le*	<appel>	*vert.*			
		eat	the.M	<apple.F>	green.M			
	b.	Ik	eet	den	roden	appel	*et*	*tu*
		manges	*la*	*<pomme>*	*verte.*			

I eat the red apple and you eat the green one. (Delbar 2022, p. 40)

In (22), *appel*, "apple," in Belgian Dutch is masculine, whereas *pomme* in French is feminine. If there were a syntactic identity relationship between the antecedent and the ellipsis, (22a) with masculine agreement in the ellipsis site would be expected to be acceptable. However, the participants in the two-alternative forced choice task preferred the feminine. In Delbar's data, the participants typically preferred the gender of the French equivalent (as in (22b) (i.e., gender agreement with the elided noun). Delbar concludes that there is no evidence for a syntactic identity relation. However, it is important to note that some studies on gender in code-switching have shown that participants may prefer the gender of the translation equivalent, influenced by the profile of the participants and the specific rate of switching within the community (see Bellamy and Parafita Couto 2022, for an overview). Consequently, as an anonymous reviewer highlighted, Delbar's findings could be relevant both to our understanding of gender agreement and the syntax of ellipsis. At the same time, the different findings for this study and previous studies on ellipsis could be due to a range of factors, including different types of ellipsis, methodological considerations (such as task effects and the number of participants), and language pairs. Delbar suggests replicating the study with Spanish–German bilinguals or using more implicit techniques such as EEG. We come back to methodological considerations in the discussion.

3.5. Information Status Particles

We have discussed several studies that included languages that use syntax and/or phonology (intonation) to encode information structure. However, in other languages, focus and topic can be expressed morphologically as well. There has been limited research on the use of these markers in code-switched utterances. However, there are some examples in the literature of these code-switches. For instance, in the English–Ewe example in (23), an Ewe topic marker is used. Moreover, the Fongbe–French example in (24) shows a novel combination involving the topic marker, ɔ́, with both a French conjunction (e.g., *donc*, "so")

and a noun (e.g., *langue*, "language"), and the doubling of Fongbe–French complementizers *ḍò -que*, "that–that" (Aboh, personal communication).

(23) English–Ewe

To	me	**dee**	mé-le	*serious*	o.	(Ameka, personal
to	me	TOP	3SG:NEG-be.at:PRES	serious	NEG	communication)

To me, he is not serious.

(24) Fongbe–French

Donc	ɔ́	nyɛ̀	mɔ́	ḍɔ́	*que*	*langue*	
so	TOP	I	see	tell	that	language	(Meechan and Poplack 1995,
ɔ́	é	dò	| |	*importante.*			p. 187)
DEF	she	be	| |	important			

So, me I see that language is important.

The Ewe and Fongbe topic markers here may be used to highlight a switch, as in Poplack's (1980) flagged switches. They may also be used to trigger another code-switch as in Clyne's (2003) Triggering Hypothesis. Ameka (2009) highlights that Kwa languages such as Akye, Akan, Ewe, Ga, Likpe, and Yoruba, while not prototypically "topic-prominent" like Chinese or "focus-prominent" like Somali, have dedicated structural positions in the clause and morphological markers to signal the information status of various components within information units. These languages can be considered "discourse configurational languages" (Kiss 1995), as they have distinct positions in the left periphery of the clause for scene-setting topics, contrastive topics, and focus. Ameka also discusses the morpho-syntactic properties of various information packaging constructions and the variations across these languages. This highlights the need for further research on the use of topic and focus markers in code-switched utterances to understand their effects.

3.6. Code-Switching between Sign Languages

There have been very few studies on code-switching between sign languages at the interfaces. An exception is Stoianov et al. (2023), who showed that reiterative code-switching between two sign languages (Cena, a young sign language used in a rural community in northeastern Brazil and Libras, the national sign language of Brazil) is used to mark information structure. In reiterative code-switching, sometimes called doubling, two signs with the same meaning (one from each language) appear one after another. In order to uncover the reasons for doubling, Stoianov et al. (2023) collected sign language data using an elicited production task. In this task, the participants described 30 short video clips of different intransitive, transitive, and ditransitive events (e.g., a woman looks at a man) to a partner, who was asked to select an image that depicted the event from three images. The task yielded 38 cases of reiterative code-switching produced by ten participants.

The results of the study showed that reiterative code-switches were particularly frequent in descriptions of reversible events, i.e., events with two animate arguments (e.g., *push, look at*) in which it is not clear who did what to whom. For instance, the results showed that the sign for WOMAN in Cena was frequently followed by the sign for WOMAN in Libras, although the reverse also occurred.

These cases of reiteration were interpreted as marking focus. The example in (25), which is produced in response to a video in which a woman gives a shirt to a man, shows a case of reiterative code-switching to mark contrastive focus. The participant initially produced the sentence in A1. As only the participant (A1) is recorded, B1's reaction is not clear. However, in A2, WOMAN appears in sentence-initial position (in Libras) and is reiterated (in Cena).

(25)

A1:	MAN WOMAN GIVE MARRIED GIVE A man gives something to a woman.	(Stoianov et al. 2023, p. 408)
B1:	[unknown]	
A2:	WOMAN (Libras) WOMAN (Cena) GIVE MAN A woman gives something to a man.	

Example (25) is thus an example of contrastive focus, in which an element is highlighted and contrasted with alternatives (in this case, other human agents who appeared in the task). These reiterative code-switches were mostly used in reversible events, and to disambiguate the sentence. Stoianov et al. argued that reiterative code-switching is used to focalize an argument (often the agent) and to disambiguate the sentence.

All in all, the paper argues that the data shed light on strategies facilitating the successful transference of potentially ambiguous information in a developing language like Cena, where conventionalized strategies have not yet taken root. Reiterative code-switches introduce a novel tool for this purpose, underscoring language users' ability to navigate ambiguity effectively by creatively employing the linguistic tools available to them. Like the studies discussed in the previous sections, this study shows the importance of information structure in accounting for code-switching. Reiterative code-switching or doubling here functions similarly to other strategies, such as intonation or word order, to focalize an element.

3.7. Summary

In summary, this section covered diverse studies on code-switching and information structure phenomena, encompassing topics like light verbs, subject pronoun-verb switches, intonation, and ellipsis. The scrutiny of light verbs entails a systematic analysis of their usage in code-switching, especially in Spanish–German (González-Vilbazo and López 2011, 2012). The study suggests that little *v,* acting as the head of the phase, dictates word order, prosodic phrasing, and information structure in code-switching constructions.

The examination of subject pronoun–verb switches started with their infrequent occurrence in corpora and their dispreferred status. However, the study by Bustin et al. (2024) underscores the role of information structure, particularly focus, in distinguishing between different theoretical models. That is, several studies on subject pronoun–verb switches argue that these switches are infrequent, and according to some models, unacceptable. However, the majority of these studies do not consider information structure. As we have shown, subject pronoun–verb switches are acceptable in certain contexts. Moreover, careful consideration of information structure (in particular, focus) allowed Bustin et al. (2024) to experimentally test the predictions of theoretical models.

We then addressed the role of intonation in conveying information structure in both unilingual and bilingual contexts. Olson and Ortega-Llebaría's (2010) study experimentally explores the impact of intonation on the perception of narrow contrastive focus in code-switched utterances. Their findings suggest that code-switching itself can function as a cue for narrow contrastive focus, especially in the absence of intonational cues.

Subsequently, the discussion shifted to ellipsis, emphasizing its connection to information structure. We underscored the limited research on ellipsis within code-switching and its potential to test theories of ellipsis. Various types of ellipsis, such as VP ellipsis, sluicing, and NP ellipsis, were explored, with a focus on studies providing support for the deletion theory and indicating a hybrid relationship between ellipsis and its antecedent.

Finally, we explored information status particles, investigating how languages express focus and topic morphologically. Once again, we highlighted the scarcity of research on these markers in code-switched utterances, with limited examples from language combinations such as English–Ewe and Fongbe–French. Stoianov et al.'s (2023) study on reiterative code-switching between two sign languages, Cena and Libras, was also discussed, revealing its use in marking information structure, particularly in disambiguating reversible events.

It is worth noting the growing interest in this area. For instance, ongoing research by Jiménez-Fernández et al. (2024) explores Topic Preposing in the grammar of Puerto Rican bilingual speakers engaging in English–Spanish code-switching, both in matrix and embedded sentences. Their hypothesis suggests that when English serves as the matrix language, preposed topics may be less accepted compared to when Spanish serves as the matrix language, due to the rigid SVO order of English in contrast with Spanish. However, their results showed that bilingual participants from Puerto Rico generally

found code-switched examples moderately acceptable regardless of the matrix language. Unilingual English clauses maintained the rigid word order of English, while unilingual Spanish examples demonstrated greater acceptability in line with the language's flexibility (Jiménez-Fernández and Miyagawa 2014; Jiménez-Fernández 2023). They are currently expanding this research to a different linguistic ecology, focusing on Spanish–English bilinguals in Gibraltar and the Virgin Islands, aiming to bridge the gap in comparative research within this domain.

Together, these studies highlight the importance of considering information structure in code-switching research, shedding light on the permissibility and acceptability of various linguistic phenomena. The findings challenge traditional frameworks and underscore the need for a more nuanced understanding of code-switching, incorporating factors such as focus, background, prosody, and community-specific practices.

4. Discussion

In this discussion, we will first examine how the interplay between information structure and code-switching enhances our understanding of multilingual grammars and language competence more generally. Second, we will address the theoretical and methodological considerations that should guide future studies in this field.

(i) The interplay between information structure and code-switching, multilingual grammars, and language competence. As we navigate through the available research on code-switching and information structure, it becomes evident that code-switching research stands at a crucial juncture, poised to make significant strides in understanding this intricate phenomenon and advancing our theoretical models. As we have shown, research on interfaces (and information structure in particular) informs code-switching research, and vice versa. The studies that we have discussed in this paper approach the topic of code-switching at the interfaces in different ways. On the one hand, there are studies that use code-switching data to inform linguistic theory, and in particular, information structure theories. For instance, González-Vilbazo and López (2012) used code-switching data to show that little v determines the grammatical properties of VP, including information structure properties. Moreover, González-Vilbazo and Ramos (2019) and Delbar (2022) used code-switching data to contribute to theories of ellipsis. Finally, Olson and Ortega-Llebaría (2010) used data on code-switching and focus to shed light on the role of intonation in bilingual grammars. On the other hand, there are studies that use information structure to test the predictions of code-switching models. For instance, Bustin et al. (2024) used theory on the information structure of subject pronouns to test the predictions from the MLF model and a minimalist approach to code-switching. These two types of approaches to code-switching and information structure are not entirely separate. For instance, the findings from Bustin et al. (2024) also contribute to theories of information structure (in particular, subject pronouns in contrastive focus), even though it was not the main objective of the study. Thus, the approaches inform each other and help advance the field. However, given the limited evidence base to date, the future of the field hinges on methodological considerations, empirical depth, and a comprehensive exploration of different theoretical perspectives. The challenges that lie ahead are both methodological and theoretical.

(ii) Theoretical and methodological considerations to guide future studies. Methodological considerations are paramount, requiring the development of more robust approaches that can overcome the challenges. One of the issues is that the empirical base is limited, often featuring isolated examples without contextualization. A lack of contextualization and attention to information structure might lead us to overstate the unacceptability of certain code-switches, such as those between subject pronouns and finite verbs. Moving forward, our understanding of the (un)acceptability of code-switches would benefit from a contextualization of examples from naturalistic data (rather than isolated examples out of context) and/or a link to a corpus of naturalistic speech (such as those found at bangortalk.org.uk or talkbank.org). Moreover, some acceptability judgment tasks might

benefit from providing utterances in an explicit context (cf. Schütze 1996), as done in some experiments (e.g., Bustin 2020; Bustin et al. 2024; Fernández-Fuertes et al. 2016). Ideally, the contexts created for experiments would be modeled on code-switching patterns observed in naturalistic data (what has been termed the field-to-cognition approach, cf. Beatty-Martínez et al. 2018; Valdés Kroff et al. 2018). It is also recommended that future research provide more detailed information on the methodology employed and, where possible, make their data available. Improved transparency regarding methodology is essential, along with an emphasis on context, as evidenced by the omission of contextual details in the majority of early studies.

As previously emphasized by Gullberg et al. (2009) and Parafita Couto et al. (2021), despite a substantial number of observations on code-switching, the support for a specific theoretical stance is limited across language combinations and both production and experimental data. There exists tension between code-switching theories, and one approach to address this is by reconciling the differences through a careful examination of patterns of variation.

To overcome the current limitations particularly regarding information structure, we must strive to build a stronger overall evidence base through comparative studies across language combinations, linguistic communities, and individual multilingual speakers, and through a wider range of research methods. First, we welcome comparative studies of a particular phenomenon across language combinations. For instance, replications of studies such as those by González-Vilbazo and Ramos (2019) on ellipsis and by Bustin et al. (2024) on subject pronoun–verb switches with different language combinations could strengthen their proposals. Moreover, research on code-switching and interfaces would benefit from more comparative studies of a particular phenomenon across linguistic communities, while keeping the language combination constant. As we discussed in Section 3, recent studies on light verbs in Spanish–English code-switching show differences between communities. The comparison across linguistic communities provides a unique opportunity to determine the interplay of linguistic and social factors.

We also argue for more studies across individual multilingual speakers, as some experimental studies have highlighted issues such as participants not being habitual code-switchers and the constraint of a limited number of participants. Future research, therefore, should include more information about the participants, their bilingual/multilingual experience, and their code-switching practices. In recent years, more tools have in fact become available to assess the participants' bilingual profile and their code-switching practices, such as the Bilingual Language Profile (Birdsong et al. 2012), the Assessment of Code-Switching Experience Survey (ACSES) (Blackburn 2013), and the Bilingual Codeswitching Profile (BCSP) (Olson 2024). More information on the participants' background and their code-switching practices would not only help in the selection of participants, but also explain variation in results within and across studies. Some recent studies, such as Bustin et al. (2024), have in fact shown variation in results across participants based on their linguistic profile (e.g., language dominance). Regarding the participants, we also argue for studies that include an examination of the participants' speech in unilingual mode. Given that these participants are bilingual, their unilingual speech may be affected by cross-linguistic influence and not follow the expectations for monolingual speech (cf. Ebert and Koronkiewicz 2018), which in some cases are crucial for the predictions of theoretical models. For instance, Bustin et al. (2024) tested the participants in both unilingual and bilingual (code-switching) modes, and only included data from participants that showed the expected pronoun use in unilingual mode in their analysis of the code-switching data. Especially given the variation across participants, it is also suggested that future research should consider increasing the number of participants.

In addition to linguistic profile, variation based on situational factors (e.g., register variation) might be relevant as well. Future studies could look at the effect of situational factors.

We also suggest that a broader array of tasks, encompassing both production and comprehension tasks in addition to acceptability judgment tasks, could enhance the scope

of investigation. Some experimental studies have raised concerns about task-related effects. Moreover, a number of recent studies on code-switching have found that findings from judgment tasks were not in line with other tasks. For instance, Bustin's (2020) findings for oral and written acceptability judgment tasks were less clear than their findings for the more implicit concurrent memory-loaded repetition task. Particularly for some of the more complex structures involving information structure, acceptability judgment tasks might not be sensitive enough to capture subtle effects. We therefore recommend including a wider range of tasks, as well as more implicit tasks (e.g., using eye-tracking or EEG).

Moving forward, we also call for more controlled experiments examining various interfaces, including discourse factors, intonation, and the role of discourse particles (cf. Carrasco Santos 2023 for a recent syntactic–prosodic–sociopragmatic approach to Spanish–English code-switching in Puerto Rico). As a first step, existing corpora could be analyzed for the use of prosody, syntax, and/or morphology in the expression of information structure in code-switched utterances. Subsequent experimental work could further investigate the interplay between different strategies in the expression of information structure and code-switching. Going back to subject pronoun–verb code-switches, most of the extant literature does not mention the prosody of the utterances, and sound files are generally not available. However, as discussed in Section 3, recent studies such as González-Vilbazo and Koronkiewicz (2016) suggest that prosody plays a role in the acceptability of these switches (even though the study was limited, in that prosody was indicated by capitalizing words). Moreover, although Bustin et al. (2024) did not include a prosodic analysis, their participants seemed to add contrastive stress in the contrastive focus condition of their concurrent memory-loaded repetition task. A good example of a study that includes both syntax and prosody is González-Vilbazo and López (2012) on light verbs. Future studies could include a more detailed prosodic analysis of the data, including repair phenomena (pauses, hesitations, repetitions, etc.) at the boundary of code-switches.

Furthermore, the evolution of theoretical perspectives hinges on a more extensive and diverse pool of code-switching data, including also code-switching (or code-blending) in bimodal bilinguals and signers, to refine and fine-tune our linguistic models. Building a stronger and more coordinated evidence base will allow us to explore the multifaceted nature of code-switching more comprehensively and enhance our theoretical models of language competence. This collective effort will enable us to refine our theories, better understand the intricate interactions at play, and ultimately advance our understanding of this complex linguistic phenomenon.

We specifically make a call for more extensive research on information structure and code-switching, exploring, for instance, various types of ellipsis and conducting studies on topic/focus markers. Additionally, there is a need for investigations across different language pairs, aiming to replicate previous findings and expand research across diverse communities. Integrating prosody, syntax, morphology, and semantics is recommended, whether studies are based on corpora or controlled production/comprehension tasks. Regarding theoretical implications, such studies constitute a promising avenue for advancing our theoretical understanding of multilingual grammars. Theories should be capable of accommodating diverse data sets, as illustrated by Bustin et al.'s discussion of how the Minimalist Program could explain their findings. As noted by an anonymous reviewer, another area for further research is the pragmatics of code-switching and its use as a rhetorical and persuasive device.

In conclusion, this review has provided a panoramic view of the existing literature on code-switching, particularly focusing on its interfaces with syntax, prosody, and discourse. Our exploration has revealed key themes, highlighted gaps in knowledge, and identified avenues for further research. In summary, a comprehensive understanding of multilingual grammars requires theoretical approaches that draw insights from diverse methodologies, including experimental studies, corpus analyses, as well as ethnographic investigations (Aboh and Parafita Couto 2024; Parafita Couto et al. 2021). As highlighted in this overview, an inclusive approach that also considers the nuances of information structure is crucial.

By integrating these various perspectives, we will not only enhance the depth and scope of our linguistic inquiries but also enrich our collective knowledge of multilingual grammars, recognizing the significance of information structure in shaping language dynamics across diverse (multilingual) linguistic landscapes. Moving forward, we anticipate that future research endeavors will build upon these insights, integrating a diverse array of methodologies and recognizing the role of information structure in multilingual grammars.

Author Contributions: Conceptualization, A.M. and M.C.P.C.; investigation, A.M. and M.C.P.C.; resources, A.M. and M.C.P.C.; writing—original draft preparation, A.M. and M.C.P.C.; writing—review and editing, A.M. and M.C.P.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Acknowledgments: We acknowledge the use of Grammarly for enhancing text readability. No AI tools were used to assist in the creation of any of the content in the manuscript. We would like to express our gratitude to the audience of the Inter-framework Information Structure workshop held in Oviedo in June 2022, as well as to the anonymous reviewers for their valuable feedback.

Conflicts of Interest: The authors declare no conflicts of interest.

Notes

1 Interfaces here does not refer to the interfaces as understood within the Y model of Minimalism (PF and LF). Instead, it refers to the syntax–discourse interface (see, for example, Erteschik-Shir 2007). While Minimalism focuses on how syntactic structures interface with the systems responsible for pronunciation (PF) and meaning (LF), emphasizing the internal modularity of the grammar, we focus on how syntactic structures are influenced by and interact with discourse context and pragmatic factors, emphasizing the external modularity of the grammar in relation to communication.

2 As pointed out by an anonymous reviewer, the "Null Theory" is attributed to Mahootian (1993). López introduced the broad/narrow distinction to characterize the work of Belazi et al. (1994) as a null theory, or constraint-free theory, presenting a re-analysis of the Functional Head Constraint (FHC) stipulating that an extended projection cannot be split. The historical debate in the field, and the issue discussed in this section of the article, is whether the Functional Head Constraint and similar models were "constraint-based" or "constraint-free." The "language feature" of the Functional Head Constraint is similar to the "language index" (Di Sciullio et al. 1986) or "language tag" (Sankoff and Poplack 1981) from prior work, making the Functional Head Constraint a code-switching-specific constraint. For an overview of these and similar issues, see MacSwan (2014).

3 The concept of "language ecologies" as discussed by Mufwene (2014, inter alia) refers to the complex environments in which languages exist and evolve. It draws an analogy with biological ecosystems, emphasizing how languages, like species, interact with each other and with their social, cultural, and historical environments.

4 Earlier work showed the presence of this construction in different language combinations. For example, Joshi (1985) provides examples of light verbs in Marathi–English code-switching and explains it through a closed-class constraint on code-switching. According to this constraint, several closed-class items like Tense, Auxiliary, and helping verbs cannot be switched between languages. Similarly, Boeschoten and Verhoeven (1985) offer instances of light verb constructions in Turkish–Dutch bilingual contexts. Additionally, Ritchie and Bhatia (1999) argue that the Minimalist Program can better account for light verbs in code-switching compared to the Matrix Language Frame model. Their research, based on Hindi–English data, demonstrates that Tense and Agreement markers cannot appear on code-switched verbs directly. To prevent the derivation from failing, a light verb with Tense and Agreement suffixes is inserted. This phenomenon is explained using the Functional Head Constraint posited by Belazi et al. (1994), which asserts that while Tense/Agreement elements select a verb phrase (VP) as their head, the Functional Head Constraint prohibits code-switching at this point. Thus, the insertion of a light verb ensures that the derivation proceeds correctly. Finally, Bandi-Rao and Dikken (2014) explore light verbs in Telugu–English code-switching and argue that a switch between a root and affixes is ungrammatical in the case of incorporation, as this creates a complex morphosyntactic head. Although this earlier work on light verbs and code-switching is very valuable, it does not relate to information structure in the same way González-Vilbazo and López's work does. Therefore, we focus on the latter in the main text.

5 González-Vilbazo and López leave open the possibility that in other language pairs, the complement of the light verb might be a root rather than a lexical verb.

[6] González-Vilbazo and López (2011) explain this based on the morphological features of little *v* in Spanish and German. Spanish has conjugation classes for verbs, and Spanish little *v* has a conjugation class feature that needs to be checked. German, however, does not have conjugation classes, and German little *v* does not have a conjugation class feature. Spanish little *v* needs to check its conjugation class feature, but it cannot be checked by a German lexical verb as German does not have conjugation classes. Therefore, the Spanish light verb, *hacer* (which has a conjugation class feature), is inserted, and the German lexical verb occurs with the default morphological suffix *-en* in German, as in *hizo nähen* in (2) (see González-Vilbazo and López 2011 for a more detailed discussion of why the light verb comes from Spanish and not German).

[7] The question remains as to what happens when *v*P is not a phase as in unaccusatives. We thank Angel Jiménez-Fernández for this suggestion.

[8] It can also be deaccented in some cases (see e.g., Frascarelli and Jiménez-Fernández 2021).

[9] As an anonymous reviewer rightly points out, examples (6) and (7) are not fully comparable, as (7) includes a clausal boundary between the main clause and its complement clause.

[10] Bustin (2020) also included oral and written judgment tasks, which are not discussed here.

[11] This concept emphasizes the importance of studying language in naturalistic contexts rather than in artificial or highly controlled experimental environments. Ecological validity ensures that the observed linguistic behaviors, patterns, and phenomena accurately reflect how language is used in everyday interactions.

References

Aboh, Enoch O., and M. Carmen Parafita Couto. 2024. Modelling multilingual ecologies beyond the L1-L2 binary. *Linguistic Approaches to Bilingualism* 14: 37–42. [CrossRef]

Aguirre, Adalberto, Jr. 1976. Acceptability Judgments of Code-Switching Phrases by Chicanos: Some Preliminary Findings. ERIC ED 129 122. Stanford: Stanford University.

Ameka, Felix. 2009. Information packaging constructions in Kwa: Microvariation and typology. In *Topics in Kwa Syntax*. Edited by Enoch Aboh and James Essegbey. Dordrecht: Springer, pp. 141–76. [CrossRef]

Balam, Osmer, M. Carmen Parafita Couto, and Hans Stadthagen-González. 2020. Bilingual verbs in three Spanish/English code-switching communities. *International Journal of Bilingualism* 24: 952–67.

Balam, Osmer, M. Carmen Parafita Couto, and Mia Chen. 2021. Being in bilingual speech: An analysis of *estar* 'be' constructions in Spanish/English code-switching. *Journal of Monolingual and Bilingual Speech* 3: 238–64. [CrossRef]

Bandi-Rao, Shoba, and Marcel den Dikken. 2014. Light switches: On v as a pivot in codeswitching, and the nature of the ban on word-internal switches. In *Grammatical Theory and Bilingual Codeswitching*. Edited by Jeff MacSwan. Cambridge: MIT Press, pp. 161–83.

Beatty-Martínez, Anne L., Jorge R. Valdés Kroff, and Paola E. Dussias. 2018. From the field to the lab: A converging methods approach to the study of codeswitching. *Languages* 3: 19. [CrossRef]

Belazi, Hedi, Edward J. Rubin, and Almeida Jacqueline Toribio. 1994. Code-switching and X-bar theory: The Functional Head Constraint. *Linguistic Inquiry* 25: 221–37.

Bellamy, Kate, and M. Carmen Parafita Couto. 2022. Gender assignment in mixed noun phrases: State of the art. In *The Acquisition of Gender*. Edited by Dalila Ayoun. Amsterdam: John Benjamins. [CrossRef]

Bellamy, Kate, Hans Stadthagen-González, and M. Carmen Parafita Couto. 2022. Él code-switches more than tú y yo: New data for the subject pronoun-verb switch constraint. *Languages* 7: 22. [CrossRef]

Bentahila, Abdelali, and Eirlys E. Davies. 1983. The syntax of Arabic-French code-switching. *Lingua* 59: 301–30. [CrossRef]

Bhatt, Rakesh, and Agnes Bolonyai. 2011. Code-switching and the optimal grammar of bilingual language use. *Bilingualism: Language and Cognition* 14: 522–46. [CrossRef]

Birdsong, David, Libby Gertken, and Mark Amengual. 2012. Bilingual Language Profile: An Easy-to-Use Instrument to Assess Bilingualism. COERLL, University of Texas at Austin. Web 20 January 2012. Available online: https://sites.la.utexas.edu/bilingual/ (accessed on 6 January 2024).

Blackburn, Angelique M. 2013. A Study of the Relationship between Codeswitching and the Bilingual Advantage: Evidence That Language Use Modulates Neural Indices of Language Processing and Cognitive Control. Unpublished doctoral dissertation, The University of Texas at San Antonio, San Antonio, TX, USA.

Boeschoten, Hendrik E., and Ludo Verhoeven. 1985. Integration niederländischer lexicalischer Elemente ins Türkische: Sprachmischung bei Immigranten der ersten und zweiten Generation. *Linguistische Berichte* 98: 346–64.

Bustin, Amy. 2020. Codeswitching in Spanish Heritage Bilinguals: Production and Perception of Switches between a Pronominal Subject and Finite Verb. Unpublished doctoral dissertation, Florida State University, Tallahassee, FL, USA.

Bustin, Amy, Antje Muntendam, and Gretchen Sunderman. 2024. Subject pronouns in Spanish-English code-switching: A test of two models. *Linguistic Approaches to Bilingualism* 14: 310–39. [CrossRef]

Carrasco Santos, Mildred. 2023. La alternancia de códigos español-inglés en Puerto Rico: Un acercamiento sintáctico-prosódico-sociopragmático. Unpublished doctoral dissertation, University of Puerto Rico, Rio Piedras, Puerto Rico.

Chomsky, Noam. 1971. Deep structure, surface structure, and semantic interpretation. In *Semantics: An Interdisciplinary Reader in Philosophy, Linguistics and Psychology*. Edited by Danny D. Steinberg and Leon A. Jakobovits. Cambridge: Cambridge University Press, pp. 183–217.

Chomsky, Noam. 1981a. Knowledge of language: Its elements and origins. *Philosophical Transactions of the Royal Society of London B, Biological Sciences* 295: 223–34.

Chomsky, Noam. 1981b. On the representation of form and function. *Linguistic Review Utrecht* 1: 3–40. [CrossRef]

Clyne, Michael. 2003. *Dynamics of Language Contact*. Cambridge: Cambridge University Press.

Delbar, Natasja. 2022. Nominal Ellipsis in French-Flemish Dutch Interclausal Code-Switched Sentences. Unpublished Master's thesis, Leiden University, Leiden, The Netherlands.

Deuchar, Margaret. 2012. Code switching. In *The Encyclopedia of Applied Linguistics*. Edited by Carol A. Chapelle. Hoboken: Wiley-Blackwell, pp. 657–64.

Deuchar, Margaret. 2020. Code-switching in linguistics: A position paper. *Languages* 5: 22. [CrossRef]

Di Sciullio, Anne-Marie, Pieter Muysken, and Rajendra Singh. 1986. Government and code-switching. *Journal of Linguistics* 22: 1–24. [CrossRef]

Ebert, Shane, and Bryan Koronkiewicz. 2018. Monolingual stimuli as a foundation for analyzing code-switching data. *Linguistic Approaches to Bilingualism* 8: 25–66. [CrossRef]

Erteschik-Shir, Nomi. 2007. *Information Structure: The Syntax-Discourse Interface*. Oxford: Oxford University Press, vol. 3.

Fernández-Fuertes, Raquel, Juana Liceras, and Anahí Alba de la Fuente. 2016. Beyond the subject DP versus the subject pronoun divide in agreement switches. In *Selected papers from the 43rd Linguistics Symposium on Romance Languages (LSRL)*. Edited by Christina Tortora, Marcel den Dikken, Ignacio Montoya and Teresa O'Neill. Amsterdam: John Benjamins, pp. 79–98. [CrossRef]

Féry, Caroline, Gisbert Fanselow, and Manfred Krifka. 2007. The notions of information structure. In *Working Papers of the SFB 632, Interdisciplinary Studies on Information Structure 6*. Potsdam: Universitätsverlag Potsdam.

Féry, Caroline, and Shinichiro Ishihara. 2022. *The Oxford Handbook of Information Structure*. Oxford: Oxford University Press.

Frascarelli, Mara, and Ángel Luis Jiménez-Fernández. 2019. Understanding partiality in pro-drop languages: An Information-Structure approach. *Syntax* 22: 162–98. [CrossRef]

Frascarelli, Mara, and Ángel Luis Jiménez-Fernández. 2021. How much room for discourse in imperative? The lens of interface on English, Italian and Spanish. *Studia Linguistica* 75: 375–434. [CrossRef]

Goldrick, Matthew, Michael Putnam, and Lara Schwarz. 2016. The future of code-mixing research: Integrating psycholinguistic and formal grammatical theories. *Bilingualism: Language and Cognition* 19: 903–6.

González-Vilbazo, Kay. 2005. Die Syntax des Code-switching. Espuglisch: Sprachwechsel an der Deutschen Schule Barcelona. Unpublished doctoral dissertation, University of Köln, Köln, Germany.

González-Vilbazo, Kay, and Bryan Koronkiewicz. 2016. Tú y yo can code-switch, nosotros cannot: Pronouns in Spanish-English code-switching. In *Spanish-English Code-Switching in the Caribbean and the US*. Edited by Rosa Guzzardo Tamargo, Catherine Mazak and M. Carmen Parafita Couto. Amsterdam: John Benjamins, pp. 237–60. [CrossRef]

González-Vilbazo, Kay, and Luis López. 2011. Some properties of light verbs in code-switching. *Lingua* 121: 832–50. [CrossRef]

González-Vilbazo, Kay, and Luis López. 2012. Little v and parametric variation. *Natural Language and Linguistic Theory* 30: 33–77. [CrossRef]

González-Vilbazo, Kay, and Sergio Ramos. 2019. Codeswitching. In *The Oxford Handbook of Ellipsis*. Edited by Jeroen van Craenenbroeck and Tanja Temmerman. Oxford: Oxford University Press. [CrossRef]

Gullberg, Marianne, Peter Indefrey, and Pieter Muysken. 2009. Research techniques for the study of code-switching. In *The Cambridge Handbook on Linguistic Code-Switching*. Edited by Barbara Bullock and Almeida Jacqueline Toribio. Cambridge: Cambridge University Press, pp. 21–39.

Gumperz, John, and Eduardo Hernández-Chávez. 1970. Cognitive aspects of bilingual communication. In *Language Use and Social Change*. Edited by Wilfred Whiteley. London: Oxford University Press, pp. 111–25.

Herring, Jon Russell, Margaret Deuchar, M. Carmen Parafita Couto, and Mónica Moro Quintanilla. 2010. 'I saw the madre': Evaluating predictions about code-switched determiner-noun sequences using Spanish-English and Welsh-English data. *International Journal of Bilingual Education and Bilingualism* 13: 553–73. [CrossRef]

Jackendoff, Ray. 1972. *Semantic Interpretation in Generative Grammar*. Cambridge: MIT Press.

Jiménez-Fernández, Ángel Luis. 2016. When discourse met null subjects. *Borealis* 5: 173–89. [CrossRef]

Jiménez-Fernández, Ángel Luis. 2023. English/Spanish relatives and their relative information structure: A view from language contact in Puerto Rico. *Catalan Journal of Linguistics* 22: 29–70. [CrossRef]

Jiménez-Fernández, Ángel Luis, and Shigeru Miyagawa. 2014. A feature-inheritance approach to root phenomena and parametric variation. *Lingua* 145: 276–302. [CrossRef]

Jiménez-Fernández, Ángel, M. Carmen Parafita Couto, and Melvin González Rivera. 2024. Topic preposing in Puerto Rican Spanish-English codeswitching and its extension to Gibraltar. Presented at 2nd Gibraltar International Conference (GIC2). Gibraltar English: Linguistic, Literary and Cultural Perspectives, Vigo, Spain, July 8–9.

Joshi, Aravind. 1982. Processing of sentences with intra-sentential code-switching. In *Coling 1982: Proceedings of the Ninth International Conference on Computational Linguistics*. Amsterdam: North-Holland Publishing Company, vol. 1, pp. 145–50.

Joshi, Aravind. 1985. Processing of sentences with intra-sentential code-switching. In *Natural Language Parsing*. Edited by David R. Dowty, Lauri Karttunen and Arnold M. Zwicky. Cambridge: Cambridge University Press, pp. 190–205.
Kiss, Katalin E. 1995. *Discourse Configurational Languages*. Oxford: Oxford University Press.
Krifka, Manfred. 2008. Basic notions of information structure. *Acta Linguistica Hungarica* 55: 243–76. [CrossRef]
Labov, William. 1971. The study of language in its social context. *Advances in the Sociology of Language* 1: 152–216.
Lambrecht, Knud. 1996. *Information Structure and Sentence form: Topic, Focus, and the Mental Representations of Discourse Referents*. Cambridge: Cambridge University Press, vol. 71.
Lipski, John. 1978. Code-switching and the problem of bilingual competence. *Aspects of Bilingualism* 250: 264.
Lipski, John. 2017. Language switching constraints: More than syntax? Data from Media Lengua. *Bilingualism: Language and Cognition* 20: 722–46. [CrossRef]
Lipski, John. 2019. Field-testing code-switching constraints: A report on a strategic languages project. *Languages* 4: 7. [CrossRef]
López, Luis. 2020. *Bilingual Grammar: Toward an Integrated Model*. Cambridge: Cambridge University Press.
MacSwan, Jeff. 1999. *A Minimalist Approach to Intra-Sentential Code-Switching*. New York: Garland.
MacSwan, Jeff. 2005. Codeswitching and generative grammar: A critique of the MLF model and some remarks on "modified minimalism". *Bilingualism: Language and Cognition* 8: 1–22. [CrossRef]
MacSwan, Jeff. 2000. The architecture of the bilingual language faculty: Evidence from intrasentential code-switching. *Bilingualism: Language and Cognition* 3: 37–54. [CrossRef]
MacSwan, Jeff. 2009. Generative approaches to code-switching. In *Cambridge Handbook of Linguistic Code-switching*. Edited by Barbara B. Bullock and Almeida Jacqueline Toribio. Cambridge: Cambridge University Press, pp. 309–35.
MacSwan, Jeff. 2014. *Grammatical Theory and Bilingual Codeswitching*. Cambridge: MIT Press.
Mahootian, Shahrzad. 1993. A Null Theory of Code-Switching. Unpublished doctoral dissertation, Northwestern University, Evanston, IL, USA.
Meechan, Marjory, and Shana Poplack. 1995. Orphan categories in bilingual discourse: Adjectivization strategies in Wolof-French and Fongbe-French. *Language Variation and Change* 7: 169–94.
Merchant, Jason. 2015. On ineffable predicates: Bilingual Greek-English code-switching under ellipsis. *Lingua* 166: 199–213. [CrossRef]
Mufwene, Salikoko. 2014. Language ecology, language evolution, and the actuation question. In *The Sociolinguistics of Grammar*. Edited by Tor Åfarli and Brit Mæhlum. Amsterdam and Philadelphia: John Benjamins, pp. 13–36.
Muysken, Pieter. 2000. *Bilingual Speech: A Typology of Code-Mixing*. Cambridge: Cambridge University Press.
Muysken, Pieter. 2013. Language contact outcomes as the result of bilingual optimization strategies. *Bilingualism: Language and Cognition* 16: 709–30.
Myers-Scotton, Carol M. 1993. Dueling Languages: Grammatical Structure in Code-Switching, 2nd ed.Oxford: Oxford University Press.
Myers-Scotton, Carol M. 2002. *Contact Linguistics: Bilingual Encounters and Grammatical Outcomes*. Oxford: Oxford University Press.
Myers-Scotton, Carol M., and Janice L. Jake. 2000. Testing the 4-M model: An introduction. *International Journal of Bilingualism* 4: 1–8. [CrossRef]
Nee, Julia. 2012. Spanish/Zapotec Code-Switching and the Theory of Ellipsis. Unpublished Bachelor's thesis, University of Chicago, Chicago, IL, USA.
Olson, Dan. 2024. The Bilingual Code-Switching Profile (BCSP): Assessing the reliability and validity of the BCSP questionnaire. *Linguistic Approaches to Bilingualism* 14: 400–33. [CrossRef]
Olson, Dan, and Marta Ortega-Llebaría. 2010. The perceptual relevance of code-switching and intonation in creating narrow focus. In *Selected Proceedings of the 4th conference on Laboratory Approaches to Spanish Phonology*. Edited by Marta Ortega-Llebaría. Somerville: Cascadilla Proceedings Project, pp. 57–68.
Parafita Couto, M. Carmen. 2005. Focus at the Interface. Ph.D. dissertation, University of Kansas, Lawrence, KS, USA.
Parafita Couto, M. Carmen, Margaret Deuchar, and Marika Fusser. 2015. How do Welsh-English bilinguals deal with conflict? Adjective-noun order resolution. In *Code-Switching at the Crossroads between Structural and Sociolinguistic Perspectives*. Edited by Gerald Stell and Kofi Yakpo. Berlin: Mouton de Gruyter, pp. 65–84.
Parafita Couto, M. Carmen, and Marianne Gullberg. 2019. Code-switching within the noun phrase: Evidence from three corpora. *International Journal of Bilingualism* 23: 695–714. [CrossRef]
Parafita Couto, M. Carmen, Miriam Greidanus Romaneli, and Kate Bellamy. 2021. Code-switching at the interface between language, culture, and cognition. *Lapurdum*. Available online: https://shs.hal.science/halshs-03280922/ (accessed on 6 January 2024).
Parafita Couto, M. Carmen, Kate Bellamy, and Felix Ameka. 2023. Theoretical approaches to multilingual switching. In *Cambridge Handbook of Third Language Acquisition and Processing*. Edited by Jennifer Cabrelli, Adel Chaouch-Orozco, Jorge González Alonso, Sergio M. Pereira Soares, Eloi Puig-Mayenco and Jason Rothman. Cambridge: Cambridge University Press, pp. 403–436. [CrossRef]
Pfaff, Carol W. 1979. Constraints on language mixing: Intrasentential code-switching and borrowing in Spanish/English. *Language* 55: 291–318. [CrossRef]
Poplack, Shana. 1980. Sometimes I'll start a sentence in Spanish Y TERMINO EN ESPAÑOL: Toward a typology of code-switching1. *Linguistics* 18: 7–8. [CrossRef]
Poplack, Shana. 2001. Code switching: Linguistic. In *International Encyclopedia of the Social & Behavioral Sciences*. Amsterdam: Elsevier, pp. 2062–65.

Ritchie, William, and Tej K. Bhatia. 1999. Code-switching, grammar, and sentence production: The problem of light verbs. In *Development of Second Language Grammars: A Generative Approach*. Edited by Elaine Klein and Gita Martohardjono. Amsterdam: John Benjamins, pp. 269–87.
Sankoff, David, and Shana Poplack. 1981. A formal grammar for code-switching. *Papers in Linguistics* 14: 3–45. [CrossRef]
Schütze, Carson. 1996. *The Empirical Base of Linguistics: Grammaticality Judgments and Linguistic Methodology*. Chicago: University of Chicago Press.
Stoianov, Diane, Anderson Almeida Silva, and Andrew Nevins. 2023. Reiterative code-switching: Argument-marking in Cena. *Sign Language Studies* 23: 386–423. [CrossRef]
Timm, Lenora A. 1975. Spanish-English code-switching: El porque y how-not-to. *Romance Philology* 28: 473–82.
Valdés Kroff, Jorge, Rosa Guzzardo Tamargo, and Giuli Dussias. 2018. Experimental contributions of eye-tracking to the understanding of comprehension processes while hearing and reading code-switches. *Linguistic Approaches to Bilingualism* 8: 98–133. [CrossRef]
van Gelderen, Elly, and Jeff MacSwan. 2008. Interface conditions and code-switching: Pronouns, lexical DPs and checking theory. *Lingua* 118: 765–76.
Vaughan-Evans, Awel, M. Carmen Parafita Couto, Bastien Boutonnet, Noriko Hoshino, Peredur Webb-Davies, Margaret Deuchar, and Guillaume Thierry. 2020. Switchmate! An electrophysiological attempt to adjudicate between competing accounts of adjective-noun code-switching. *Frontiers in Psychology* 11: 549762. [CrossRef]
Weinreich, Uriel. 1953. *Languages in Contact*. Mouton: The Hague.
Winkler, Susanne. 2022. Ellipsis and information structure. In *The Oxford Handbook of Information Structure*. Edited by Caroline Féry and Shinichiro Ishihara. Oxford: Oxford University Press. [CrossRef]
Woolford, Ellen. 1983. Bilingual code-switching and syntactic theory. *Linguistic Inquiry* 14: 520–36.
Zubizarreta, Maria Luisa. 1998. *Prosody, Focus and Word Order*. Linguistic Inquiry Monographs 33. Cambridge: The MIT Press.

languages

MDPI

Article

The Syntax of Speech Acts: Deictic Inversion as an Evidential Strategy in English

Ana Ojea

Department of English, French and German Philology, University of Oviedo, 33011 Oviedo, Asturias, Spain; aojea@uniovi.es

Abstract: This paper presents empirical evidence to support the so-called syntactization of discourse, that is, the projection of relevant pragmatic features in the narrow syntax. In particular, it analyses *deictic inversion* in English, a construction which is used by the speaker to point at a proximal or distal location and bring the addressee's attention to an entity related to that location (e.g., *Here comes the bus*). It offers a novel account of this construction, which takes it to be an evidential strategy in a language that does not have standard evidential markers; this evidential status explains its main differences with locative inversion, a construction with which it is pragmatically and structurally related. Deictic inversion therefore receives a natural explanation in a framework that maps syntax with the speech act and introduces in the derivation pragmatic information about the participants in the communicative exchange and about the source of the information for the proposition asserted.

Keywords: speech act; evidentiality; deictic inversion; locative inversion; syntax-pragmatics interface

Citation: Ojea, Ana. 2024. The Syntax of Speech Acts: Deictic Inversion as an Evidential Strategy in English. *Languages* 9: 183. https://doi.org/10.3390/languages9050183

Academic Editor: Claudia Felser

Received: 21 December 2023
Revised: 1 May 2024
Accepted: 7 May 2024
Published: 17 May 2024

1. Introduction

In recent years, generative grammarians have amply discussed whether pragmatic features should be configurationally represented, and, if so, which pragmatic information should be syntactically encoded in terms of specific categories and structural relations. In this respect, there exists general consensus that the different types of topics and foci must be syntactically represented, along with the type of information they convey.

Together with an articulation of the left/right periphery to represent information structure, there has been another productive line of research that explores the mapping between syntax and the speech act, that is, the need to introduce in the narrow syntax the discourse participants speaker/addressee, along with the notions of commitment, evidentiality and evaluation in which they are involved. Investigations along these lines have successfully shown the effects of various aspects of pragmatic prominence in the syntactic structure and have explained in a principled way a number of phenomena whose grammatical properties are crucially determined by the discursive status of the proposition (cf. among others, Ross 1970; Cinque 1999; Smith 2000; Speas and Tenny 2003; Speas 2004; Haegeman and Hill 2013; Haegeman 2014; Miyagawa 2022; Krifka 2023).

This paper goes in this direction, and here I propose a novel analysis of the so-called deictic inversion (DI) in English which hinges on an explicit codification of the relevant features active in the communication exchange and incorporates some of the insightful observations about the form and function of the construction made in Lakoff (1987).

Deictic inversion is used by the speaker to point at a proximal or distal location and bring the addressee's attention to an entity related to that location. It therefore requires a perceptual field shared by both the speaker and the addressee, and this, as Green (1982, p. 130) has claimed, is what makes it basically an oral language construction:

(1) Here comes the bus.
(2) Here comes Max with his new girlfriend.
(3) There goes Mary.
(4) There goes a beautiful car.

As I will argue below, the structural properties of examples such as (1)–(4) (taken from Lakoff 1987) can be accounted for in terms of valuation of a pragmatic feature in the speech act projections of the sentence by a lexical verb, something that not only explains the full inversion of the verb with the subject in DI, but also certain restrictions in the form and reading of the verbal form and in the distribution of the construction.

The paper is organised as follows. Section 2 summarises the main arguments that have been used in the relevant literature to justify the syntactic projection of the discourse participants (speaker and addressee) and of the notion of evidentiality, both of which are essential for our understanding of a number of grammatical phenomena, among them deictic inversion. Section 3 presents an analysis of DI which formalises the role played in the derivation by a discourse feature of evidentiality and explains the main structural properties of the construction in a principled way; this analysis will also account for the differences between DI and locative inversion (*At the gathering arrived some unexpected visitors*), two constructions typically grouped together in grammatical descriptions. Section 4 offers some conclusions.

2. The Syntax of Speech Acts

Issues related to speech acts (i.e., *how to do things with words*, to use Austin's (1962) seminal formulation) pertain to the actual use of language in communication and therefore have generally been treated as merely pragmatic. Nonetheless, a number of influential studies in the last few decades have led to what has been termed the "syntactization of discourse" (cf. Haegeman and Hill 2013), that is, the recognition that there must be a representation of the speaker and the addressee in the syntax, alongside some other structural layers that mediate between the communicative act and the meaning of the utterance.[1]

2.1. Speech Act Projections

As early as in 1970, Ross proposed that not only performative sentences in the sense of Austin (1962) (i.e., declarations, directives, commissives. . .), but also assertive declarative sentences should be derived from deep structures with a covert superordinate structure that contains a performative verb, the speaker and the addressee. Ross (1970) also observed that discursive relations are constrained by the same kind of hierarchical rules that constrain syntactic relations, a point made in Oswalt (1986) and Willett (1988) as well; similarly, Cinque (1999), in his influential work on the cartography of clausal functional projections, notes that those morphemes that express the source of information and evaluation of the sentence show striking crosslinguistic regularities in their position within a word.

Nevertheless, even if hierarchical relations in discourse significantly resemble those of the computational component, this is not enough to propose that discourse features encoding the speaker/addressee and their point of view in the proposition must be projected in the narrow syntax. One should also find robust empirical data which clearly show an interaction between the communicative act and some syntactic operation, that is, constructions in which the grammatical form crucially depends on the discursive properties of the proposition.

This seems to be the case of evidential morphemes in a number of languages, i.e., the morphemes that mark the speaker's source for the information being reported in the utterance (see the next section); or logophoric pronouns in some African and native American languages, which refer to some individual whose point of view is being represented in the sentence (for details about logophoricity, see Sells 1987, Speas 2004, Miyagawa 2022, and references therein). Speas (2004) convincingly argues that the distribution of these grammatical elements can only be accounted for if one adopts a framework in which there are some syntactic projections that bear pragmatic features for the notions of speech act, evaluation, evidentiality and epistemicity. Along the same lines, Speas and Tenny (2003, p. 17) list a number of constructions whose description requires reference to some sentient individual, other than speaker or addressee, noting that there are systematic restrictions in all of them

that would be surprising if the discourse-related properties of these constructions were purely pragmatic.

As for the pragmatic features that should be included in the syntactic representation, Speas (2004) adopts Cinque (1999)'s projections for Speech Act Mood, Evaluative Mood, Evidential Mood and Epistemological Mode, hierarchically organized above CP (or above ForceP, if one assumes the fine structure of the left periphery laid out in Rizzi 1997). She articulates these discourse categories under a Speech Act projection (SAP), and associates each of them with an implicit argument in their specifier position; these arguments represent the speaker, the evaluator, the witness and the perceiver, respectively:

(5) Speas (2004)'s Speech Act Phrase:
$[_{SAP}$ Speaker SA $[_{EvalP}$ Evaluator Eval $[_{EvidP}$ Witness Evid $[_{EpisP}$ Perceiver Epis $[_{CP}$...

As can be observed, the structure in (5) distinguishes the notion of evidentiality from the closely connected notions of evaluative mood and epistemological modality, all of which measure the information status of the sentence and share two salient properties: they involve a source of evaluation/reliability for the sentence and offer a scalar measure of the information status of that sentence (vid., Rooryck 2001). Actually, many analyses do not project them as separate categories (but see Section 2.2 below). For example, Speas and Tenny (2003) collapse this information into a Sentience Phrase, a projection whose arguments are the seat of knowledge (i.e., the sentient mind that can evaluate or comment on the truth value of the proposition) and evidence (i.e., the type of evidence available for evaluating that truth). This Sentience Phrase is dominated by the Speech Act Phrase, which includes the speaker and the hearer; in an unmarked statement, the speaker is the seat of knowledge, and in a question, the seat of knowledge is the hearer.

The analyses in Speas (2004) and Speas and Tenny (2003) opened the door to a growing body of research devoted to the structure of speech acts (Sigurðsson 2004; Zanuttini 2008; Coniglio and Zegrean 2012; Miyagawa 2012, 2017, 2022; Woods 2016; Wiltschko and Heim 2016; Portner et al. 2019; Wiltschko 2021; Krifka 2001, 2015, 2023). For example, in his influential works, Krifka holds that a speech act obtains when a proposition joins to three structurally and functionally distinct layers that codify its illocutionary force: a Judgement Phrase, which represents subjective epistemic and evidential attitudes; a Commitment Phrase, which represents the social commitment related to assertion; and an Act Phrase, which represents the relation to the common ground of the conversation and distinguishes assertions from questions. The participants in the speech act are not explicitly represented in Krifka's model (see Krifka 2023 and references therein).[2]

Recently, Miyagawa (2022) has returned to the idea that the speaker and the addressee should be represented in the syntactic structure and proposes a top layer, modelled on the Speech Act Phrase in Speas and Tenny (2003), with a shell-projection where SAP introduces the Speaker and saP the Addressee.[3] Miyagawa (2022) also claims that the syntactic structure of a sentence is essentially partitioned into an expressive component above CP, which is about the performative act and does not contribute to the truth-value of the utterance; and a propositional component, reserved for elements that form the proposition and concern truth conditions. He integrates Krifka's JudgeP in the propositional component (i.e., in the C-system), arguing that some linguistic elements that belong to this JudgeP contribute to the truth-value of the proposition (for example, certain modals and adverbs), and just leaves SAP and CommitP as the syntactic bases for the speech act:

(6) Miyagawa's SAP:

[SAP [CommitP	[CP [TP…
Expressive component	Propositional component

In what follows, I will assume Miyagawa's analysis of speech acts in terms of an expressive component and a propositional component, but I will single out evidentiality as an independent category (EvidP) in the expressive component, in the spirit of Cinque (1999) and Speas (2004).[4] To simplify the representations, I only project in the expressive component the categories SAP and EvidP, both essential for the analysis of DI; I therefore omit CommitP, which in this construction would have a head marking that the speaker makes a public commitment to the proposition. The relevant (simplified) structure for speech acts that I adopt here will then be as in (7):

(7) $[_{SAP}$ Speaker SA $[_{saP}$ Addressee sa $[_{EvidP}$ Evidence Ev $[_{CP}$ $[_{TP}$ . . .

As I will show, the structure in (7) serves to offer a principled account of a construction such as DI, where the form of the sentence is clearly determined by its communicative function. Given the relevance of the notion of evidentiality for the proposal, the next section further discusses its status as an illocutionary functional category above CP.

2.2. The Syntactic Projection of Evidentiality

According to Jacobsen (1986), the term *evidentiality* was introduced into linguistics in a posthumously published grammar of Kwakiutl compiled by Franz Boas in 1947 (Boas 1947), and it was brought into common usage by Jakobson (1957). Since then, the topic has been dealt with from a wide variety of perspectives, ranging from typological studies to cognitive linguistics, grammatical description, and pragmatics.

Evidentiality refers to the grammatical expression of the information source for the content of the proposition and, as such, it serves to put that proposition in perspective. As was mentioned above, the relationship between evidentiality (which marks the source of the information), and epistemic modality (which marks the degree of reliability of that information) is not always easy to demarcate; for this reason, evidentiality has sometimes been treated as a subcategory of epistemic modality, under the view that the degree of commitment to the information depends on the information source, since this will be more reliable if the evidence is direct than if it is indirect; see, among others, Chafe and Nichols (1986), Palmer (1986) and Izvorski (1997). Krifka (2023), for example, encodes both, epistemic and evidential attitudes, in the single category Judgement Phrase; unlike Cinque (1999), Krifka does not make a structural distinction between evidential (*reportedly*, *allegedly*. . .), and epistemic adverbials (*probably*, *possibly*. . .) either, and projects the two of them as modifiers of JudgeP —one expressing the source of the judgement and the other the strength.

Other approaches treat the notions of evidentiality and epistemic modality as distinct but closely related. For example, Boye (2012, pp. 2–3) views evidentiality and epistemic modality as two subcategories of an epistemic domain: evidentiality will provide epistemic justification for the truth of the proposition (i.e., source of information, evidence or justification), whereas epistemic modality will provide epistemic support for it (i.e., degree of certainty and degree of commitment). Likewise, González et al. (2017) describe how certain lexical and grammatical resources can have both evidential and epistemic uses.

Finally, there are analyses that separate the evidential marking of the source of information from the speaker's degree of confidence about the truth of the propositional content, on the basis of the differences that exist between the two. For example, de Haan (1999, 2005) argues that evidentials and epistemic modals differ in their lexical origins, and they also differ semantically: epistemic modality evaluates the evidence, whereas evidentiality asserts that evidence; moreover, fully-grammaticalized evidentials cannot occur within the scope of negation, unlike epistemic modal elements. Aikhenvald (2004, 2015) also points out that, even though evidentials can have epistemic extensions (relating to the degree of the speaker's certainty concerning the statement), this does not need to be always the case, which for her means that evidentiality can be considered a category in its own right and not a subcategory of a specific type of modality (cf. Aikhenvald 2004; Nuyts 2005, among others). And Faller (2002, 2006) shows that, in languages such as Quechua, evidentiality and epistemic modality are expressed by clearly distinct sets of linguistic

markers, which supports the development of a theory for each notion independently; she admits, though, that this does not preclude the possibility that specific linguistic markers may combine both.

To try and offer a conclusive answer to how evidentiality must be conceived in relation to epistemicity is well beyond the aim of this paper, but there seems to be sufficient ground to hold that they can be projected separately, and this is the turn that I will take here. It is then necessary to see whether the type of semantics involved in evidentials is illocutionary or propositional, since this is crucial to determine whether EvidP should be projected in the expressive component or in the propositional component (put differently, outside or inside the CP layer). It should be noted in this connection that even in those approaches that relate evidentiality to epistemicity, a distinction is customarily drawn between illocutionary evidentiality and epistemic evidentiality (see, among others, Izvorski 1997; Faller 2002, 2006; Matthewson et al. 2007; Murray 2010, 2021 and Demonte and Fernández-Soriano 2022). Assuming this divide, illocutionary evidentials are treated as functions from speech acts to speech acts, since they may modify the sincerity conditions of the act they apply to, but do not add to the propositional content of a sentence; thus, their contribution is not directly challengeable or up for negotiation (cf. Faller 2002, p. 231); on the contrary, epistemic evidentials may contribute to the propositional content and can be treated as epistemic modals with an evidential presupposition.[5] Illocutionary evidentials will then contribute to the illocutionary or speech act content, while epistemic evidentials contribute to the propositional content. This distinction, particularly relevant in the case of indirect evidentials, has been tested on a set of diagnostics which basically check if embedding is possible, whether the evidential contribution can be challenged or not, what the relevant scope of evidentials is with respect to tense and modals, and how evidentials interact with questions (see Murray 2010). As will be shown at length below, DI—which involves direct evidentiality—clearly patterns with illocutionary evidentials under these tests: embedding is impossible, the content of the evidential operator cannot be challenged, the evidential cannot be in the scope of tense or modality, and there is no interaction with questions.

Therefore, the category EvidP in (7) stands for illocutionary evidentiality, and I assume that it contains features relevant for the speech act, not for the truth-conditional meaning of the sentence (i.e., it belongs to the expressive component). With regard to the particular features that head the projection, obviously only a short number out of the potentially infinite set of sources of evidence are grammaticized in evidential paradigms (cf., Speas 2004, p. 257). The main distinction here is between direct (i.e., attested) and indirect (i.e., inferred, or reportative) evidence (see Willett 1988, p. 57); Aikhenvald (2015, p. 240) reports the following as the recurrent meanings found in the evidential systems of human languages:

(8)
- (I) Visual: evidence acquired through seeing;
- (II) Sensory: evidence through hearing, typically extended to smell and taste, and sometimes also touch;
- (III) Inference: visible or tangible evidence, or visible results;
- (IV) Assumption: based on reasoning and conjecture (and not on visible results);
- (V) Reported: reported information with no reference to who it was reported by;
- (VI) Quotative: reported information with an overt reference to the quoted source.

One could then assume that the head of EvidP projects a [direct] or [indirect] discourse feature, together with the following specifications:

(9) Evid
[Direct] → {[visual] [sensory]}
[Indirect] → {[inference] [assumption] [reported] [quotative]}

It is estimated that around one quarter of the world's languages have an evidential system (Aikhenvald 2004, p. 30). In these languages, evidential meanings are grammatically realized as autonomous particles or as (lexical or covert) morphemes fused with some other syntactically projected feature, normally tense or aspect (cf., Palmer 1986; Willett 1988; Aikhenvald 2015).[6] An example of a language with autonomous ev-

identiality is Nheêngatú or Língua Geral, a Tupí-Guaraní lingua franca of north-west Amazonia. In this language, if the speakers want to assert something for which they just have indirect reported evidence, they use the autonomous evidential marker *paá*:[7]

(10) u-sú u-piniatika paá
3SG-go 3SG-fish REP
'He went fishing (they say/I was told)'.

As for morphological evidentiality, we find examples in many Amerindian languages, as is the case of Jarawara, an Arawá language from Brazil, which has a direct (first-hand) and an indirect (non-firsthand) information source whose expression is fused with tense. For example, if a man is woken up by a dog (and has seen and/or heard it), he would use the direct sensory evidential morpheme *-are* fused with the immediate past:

(11) owa na-tafi-are-ka
1 SG.O CAUS-wake-IMMPST.EYEWIT.M-DECL.M
'It did wake me (I saw it or heard it)'.

English, as most Indo-European languages, lacks evidential markers of this sort and generally expresses evidentiality lexically, through adverbs (*reportedly*, *allegedly*...), complex prepositions (*according to*, *as claimed by*...) or complex sentences headed by perception verbs (*I hear*, *I can see*...), perception semi-copulas (*looks like*, *sound*...) or verbs of speaking (*be said to*, *they say*...); see Mélac (2022, p. 234). To these evidential strategies, I would like to add a syntactic construction which, to my knowledge, has not been explicitly approached this way in the relevant literature: deictic inversion. As I will show below, deictic inversion in English has some defining properties that find a natural explanation in a model which represents the syntax of speech acts and includes (a) the participants in the communicative exchange and (b) an evidential feature that encodes the speaker's qualification of the proposition in terms of the type of evidence available for evaluating its truth. In the next section, I turn to this task.

3. A Formal Analysis of Deictic Inversion in English

English has an unmarked subject–verb order in declarative sentences, with the subject placed in front of the verb for formal reasons (i.e., it is an agreement-prominent language in the sense of Jiménez-Fernández and Miyagawa 2014). In a number of constructions, though, this basic word order is altered, and the subject occurs in a post-verbal position (following the lexical verb) while some other constituent is fronted. This so-called full inversion of the subject obtains for reasons which basically have to do with information packaging, the general condition being that the postponed subject, which becomes the informational focus, is less familiar informationally than the fronted constituent (see Birner 1996 and references therein); (12)–(14) are examples of constructions which involve full inversion in English:[8]

(12) 'Leave me alone!' shouts Harry. (quotative inversion)
(13) In walked the cat. (directional inversion)
(14) Behind him came Eton Lad, who fluttered. (locative inversion)

Full inversion for discursive reasons is also found in deictic inversion (DI) structures, such as (1)–(4) repeated here as (15)–(18), which are used by the speaker to bring the addressee's attention to an entity related to a proximal or distal location:

(15) Here comes the bus.
(16) Here comes Max with his new girlfriend.
(17) There goes Mary.
(18) There goes a beautiful car.

DI can be headed by the unaccusative verbs *come* and *go*, as in (15)–(18), or by copula *be*:[9]

(19) There is Harry with his red hat on.

In this paper, I focus on the cases where DI is headed by a predicative verb, as in (15)–(18) above (i.e., the so-called *perceptual deictic (sub)construction* in Lakoff 1987, p. 482), and will only refer to the general construction with *be* when it bears on aspects which are relevant for the description.

Superficially, DI has much in common with standard locative inversion (LI). In both cases, the sentence is conceived as a non-predicative assertion of a state of affairs where the grammatical subject is a participant involved in that event (not the entity the proposition is about) and receives the informational focus. And both, DI and LI, have a locative constituent in initial position and are headed by a copula or an unaccusative verb (i.e., a verb that lacks an external argument and is informationally light); as examples (14) and (15) show, unaccusatives *come* and *go* may head both constructions.

Despite these similarities, DI cannot just be approached as a subtype of locative inversion since it has a discursive status which is different from that of LI. If one assumes that pragmatic information is syntactically encoded, the derivation of the two constructions is then predicted to be different as well, something which, in turn, will explain their structural differences. To show this, I will first discuss the derivation of LI to then compare it with DI and offer an analysis of the latter which shows the crucial role that a discursive feature of evidentiality has in its final form and distribution.

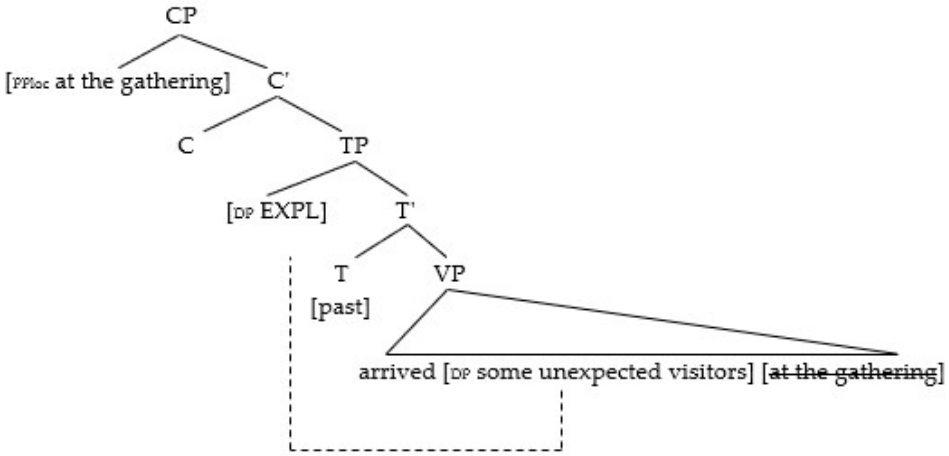
3.1. DI As an Evidential Strategy in a Non-Evidential Language

LI is a stylistic mechanism which has a presentational function. Structurally, it places a locative constituent in initial position to then (re)introduce the subject in the part of the scene that the fronted locative refers to (see, among others, Bresnan 1994; Levin and Rappaport Hovav 1995; Birner 1996; Dorgeloh 1997 and references therein); therefore, any constituent which serves to locate the predicate may be fronted for this purpose. As for the verb, as mentioned above, the only condition is that it is not agentive and does not contribute new information to the discourse (i.e., it must be informationally light). The construction can thus be headed by copula *be* (20), unaccusative verbs of inherently directed motion, appearance and existence (21)–(23), and also unergative verbs that have been pragmatically emptied of their agentive meaning, that is, "unaccusativized" in the sense of Torrego (1989), as in (24):[10]

(20) In the vase are some flowers.
(21) At the gathering arrived some unexpected visitors.
(22) On the stage appeared a hideous creature.
(23) Near his house lies a buried treasure chest.
(24) Among the guests was sitting my friend Rose.

The (simplified) derivation of a LI structure such as (22) will be as follows (see, among others, Postal 1977, 2004; Hoekstra and Mulder 1990; Chomsky 2008; Bruening 2010; Ojea 2019, 2020):[11]

(25)



As (25) shows, in the derivation of LI a locative phrase is fronted into CP, and the rest of the constituents remain in their underlying position. This is because T values its formal features on V via agreement—as is unmarkedly the case in English when the verb is lexical—and therefore, the lexical verb does not need to leave the VP domain. As for the subject (*some unexpected visitors*), it is E-merged VP-internally (i.e., it is an internal argument, since the verb is unaccusative) and remains there because the structural subject position Spec-TP is occupied by a covert expletive; this covert expletive receives empirical justification on the grounds of the existence of sentences where it is overtly realized:[12]

(26) On the stage, there appeared a hideous creature.
(27) Near his house, there lies a buried treasure chest.
(28) At the gathering, there arrived some unexpected visitors.

The covert expletive hence satisfies the formal EPP requirement of T. As is well known, though, the expletive only has a partial set of phi-features (specifically, a person feature), and therefore, T must probe the DP subject *some unexpected visitors* to value the rest of its features on it, thus inducing verb–subject agreement. As for the nominative case feature of the subject, it gets valued via coindexing with the expletive in Spec-TP, with which it forms an A-chain.

Superficially, DI may look like a sub-type of LI where the verb that heads the construction is lexically restricted (only *be* and unaccusatives *come* and *go)* and so is the locative constituent that is fronted (just *here* or *there*). Significantly, both the verb and the adverb have a locative deictic component which is measured with respect to the speaker: *come* expresses motion towards the speaker, whereas *go* expresses motion away from the speaker; *here* points at a proximal location with respect to the speaker, and *there* at a distal location with respect to the speaker. Moreover, in discourse, speakers use DI with a particularly complex intention which is not there in standard LI and involves coordinated acts and effects on three cognitive dimensions: speaking, visual perception and the construction of spatial mental models on the part of the addressee (cf., Webelhuth 2011, p. 91): the speaker brings the addressee's attention to an entity (related to a proximal or distal location), which thus constitutes the informational focus of the proposition. In other words, DI is used as an evidential strategy when the speaker commits to the truth of a proposition relying on direct (visual) evidence and wants to make the addressee aware of this.

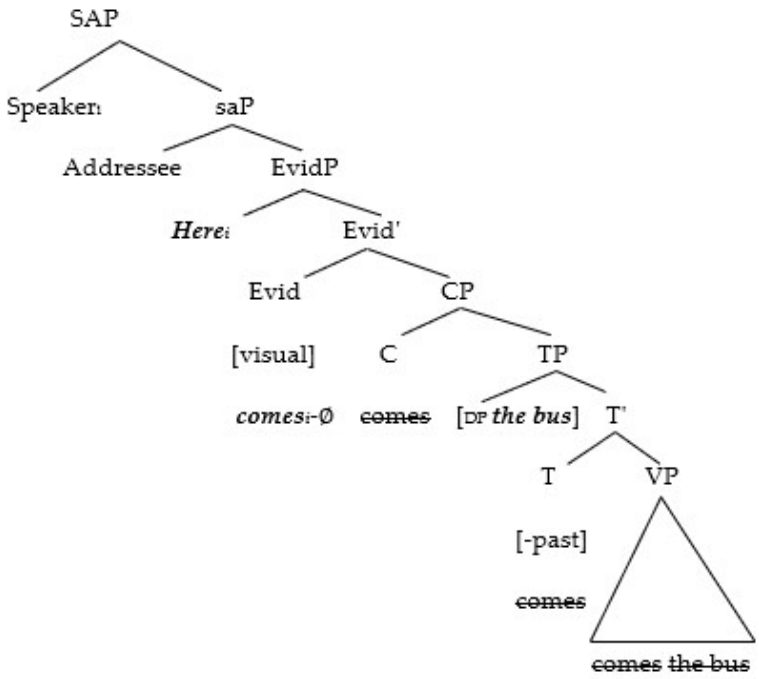
My proposal here is that it is this evidential status that determines the structural properties of the construction. To show this, I assume an analysis of speech acts along the lines discussed in Section 2, adopting the (simplified) structure in (7) (repeated here as (29)),

modelled on Miyagawa (2022), but with evidentiality projected as a category in its own right in the expressive component:

(29) $[_{SAP}$ Speaker SA $[_{saP}$ Addressee sa $[_{EvidP}$ Evidence Ev $[_{CP}$ $[_{TP}$. . .

The (simplified) derivation of a sentence such as (21) (*Here comes the bus*) will then be as follows:

(30)



As (30) shows, contrary to the case of LI, the derivation of DI crucially relies on the discursive categories in the expressive component, and both the verb and the locative adverbial—coindexed with the speaker—are placed in EvidP. In DI, the information source for the content of the proposition is always direct evidence on the part of the speaker, and EvidP is headed by the δ-feature [visual], encoded as a covert morpheme (one of the possibilities found in the evidential paradigm crosslinguistically); the head feature may be [sensory] instead of [visual], as in (31) and (32), when the source of the information is non-visual sensory experience (see Lakoff 1987, p. 484):

(31) Here comes the beep.
(Auditory evidence: when you hear the warning click of the alarm clock)

(32) Here comes the pain in my knee.
(Physical evidence: when you feel a twinge before the pain appears)

As mentioned, the derivation of DI involves the expressive component, and the main structural differences between DI and LI have to do with the placement of the locative adverbial, the position of the subject and the eventual placement of the verb in the structure.

In DI, the deictic adverbial *here*/*there* (coindexed with the speaker) is E-merged in the evidential projection to mark the visual reference point as proximal or distal; the adverbial may therefore coexist with the expression of some other locative complement in the VP:

(33) Here comes a bus into the terminal.

This contrasts with LI, where the fronted adverbial is I-merged in the CP projection and therefore leaves a copy in its underlying position, thus preventing the projection of another constituent of the same type:

(34) *At the gathering arrived some unexpected visitors there.

As for the position of the subject, whereas in LI the structural subject position is occupied by an expletive, this is not the case in DI, where the presence of expletives is ruled out (compare (35), with an unstressed *there* (/ðə(r)/) in subject position, with (26)–(28) above):

(35) *Here there comes the bus.

Therefore, the DP subject in DI structures is targeted into TP to satisfy the EPP, and it also values its own case feature there. Since the subject sits in the canonical Spec-TP position, no definiteness effect will be at play (*Here comes a bus/the bus/Max)* and, as expected, there will be agreement of the DP with the verb (examples taken from Kay and Michaelis 2017, p. 19):

(36) a. There goes John's old tutor.
b. There go two boys who just turned twenty...

If the DP subject is in Spec-TP, we would expect to find the same type of DPs here that we may find in any other assertive sentence, including pronominal DPs. These are nevertheless forbidden in the construction:

(37) *There comes he.

Note, though, that this impediment to have a pronominal subject postverbally affects not only DI but all of the constructions which involve full inversion in English; compare (38)–(40) with (12)–(14):

(38) *'Leave me alone!' shouts he.
(39) *In walked it.
(40) *At the gathering arrived they.

In all of these constructions, when the structure is transferred to the conceptual-intentional system, the subject follows the lexical verb and must be interpreted as the informational focus of the sentence. This suggests that the impossibility to have a postverbal pronominal subject in sentences such as (37)–(40) does not have to do with a formal restriction, but with a pragmatic constraint related to information requirements: given that the DP subject constitutes the informational focus, it must convey new—or at least less familiar—information than the other constituents (cf., Birner 1996's Relative Familiarity Constraint). This is what rules out anaphoric pronouns, which, by definition, refer back to entities already in the common ground. If the pronoun contributed new or contrastive information (i.e., if it had a heavy stress and a focal reading) it could actually be a possible subject in DI, and this is attested by some native speakers who claim that, if the sentence in (37) were inserted in any of the dialogues below, it would be acceptable (stress indicated with capitals):[13]

(41) The only person who could save us now is Bob.
Oh, look! Here comes **HE!**
(42) We really need Josh and Katie to get here, right now!
Oh look, there comes **HE**, at least, though I still don't see her anywhere.

It is also significant that the pronominal subjects in (41) and (42) are in the nominative case, something which provides additional evidence for the placement of the subject in the Spec-TP position.[14]

Finally, the eventual placement of the verb in the head of EvidP has to do with the role of the verbal predicate in DI. Evidentials behave like indexicals (cf. Kaplan 1989), and direct evidentiality, in particular, is speaker-anchored. In English, there is no lexical morpheme which may mark this indexical relationship with the speaker and, therefore, the construction resorts to two unaccusative predicates, *come* and *go*, which include in their meaning a component PATH measured towards/away from the speaker and locate the speaker at the starting/end point of the motion event. The unaccusative predicate coindexed with the speaker is targeted into EvidP in the expressive component to value the evidential feature there, so that the sequence can be successfully transferred to the conceptual-intentional system. This means that the verb leaves the VP domain, contrary to the case of LI (see (25)), and this movement must satisfy the general considerations of simplicity and efficient design, i.e., it must take place in the most economical way under locality conditions. The implication is that the verb must move into Evid in a head-to-head fashion and thus values the formal features of T on its way to the speech act projection. Eventually, then, tense and evidentiality are fused, as is also the case in evidential systems, where evidential markers tend to appear fused with some syntactically projected grammatical feature (normally tense or aspect).

This derivation implies that T does not value its features under agreement with *come* and *go*, contrary to what is standardly the case with lexical verbs in present-day English; in other words, in DI, *come* and *go* behave as if they were auxiliary verbs. One should bear in mind, in this respect, that the overt movement of *come/go* in the construction is necessary for convergence with the external intentional system, that is, it is a marked operation where interface economy competes with computational economy, forcing a costly derivation (on interface economy, see Reinhart 2006). Furthermore, *come* and *go* group in DI with copula *be*, and they just differ in that the unaccusative predicates have a deictic locative component which is not there in the meaning of the auxiliary, and serves to mark more explicitly the spatial relationship between the speaker and the entity signalled (e.g., *There is/goes Harry with his red hat on*). Therefore, the fact that copula *be* (a real auxiliary) and *come* and *go* behave alike syntactically in DI may just be a natural consequence of the little semantic import of the latter, which makes them auxiliary-like in the construction.[15] Note as well that *come* and *go* have a functional behavior in pseudo-coordinations (e.g., *What has John gone and done all day?*; see de Vos (2004) for details), which means that they can be semantically bleached in other constructions too.[16]

To summarize so far: DI can be analysed as a syntactic strategy which marks evidentiality in a language that is not evidential in the strict (morphological) sense. In it, the verb is eventually placed outside CP, in the head of EvidP in the expressive component, and there it values a discourse feature that encodes direct evidentiality—the information source for the content of the proposition is visual, or at least sensory, evidence. The deictic adverbial *here* or *there* in the specifier of that projection signals whether the visual reference point is proximate or distal to the speaker.

As I will show next, an analysis along these lines not only accounts for the discursive and formal properties of DI just mentioned but also explains the main differences between DI and LI, which, as expected, basically follow from the different illocutionary value of the two constructions.

3.2. Empirical Predictions of the Analysis

The analysis of DI in (30) explains one of the aspects which most notably distinguishes this construction from its non-inverted counterpart and from similar structures such as LI: the temporal interpretation of the verbal form. As discussed above, the evidential feature in the expressive component targets the lexical verb, which values tense on its way up. This means that the grammatical feature tense will eventually be fused with the discourse feature of evidentiality that marks that the speaker has visual or sensory evidence of the facts. Accordingly, if the verbal form is present (see all of the examples used so far), it will have the interpretation that the speaker has direct evidence for the proposition at the

moment of speech. Consequently, the simple present does not have here the imperfective generic interpretation which is the unmarked reading of a "true" simple present in English (i.e., it will be a *present of evidentiality*). This is quite evident if we compare DI with its non-inverted counterpart:

(43) Here comes a bus now/* regularly.
(*vs.* A bus comes here * now/regularly.)

A similar situation is found in the past. Even though most instances of DI are in the present tense, examples of DI where the verbal predicate has past morphology are also possible, as Kay and Michaelis (2017, p. 21) show:

(44) Here came the waitress. She had on a mini-skirt, high heels, see-through blouse with padded brassiere.
(45) So I looked, and here came a white horse!
(46) Here came the Princess, and as she passed hats were lifted.
(47) There went Dr. and Mrs. Sorabjee, leaving little Amy alone at their table.

The past in these sentences must also be understood as a *past of evidentiality*, since it marks that the speaker had direct evidence of a situation which was ongoing at the reference time of the narration. This reading, as in the case of the present above, contrasts with the standard reading of the simple past in English, which unmarkedly places the event as anterior to the time of the assertion; note in this respect that the sentences in (44)–(47) are not compatible with an adverbial such as *yesterday*, which marks anteriority:

(48) *Yesterday, here came the waitress. . .
(49) *Yesterday I looked, and here came a white horse!
(50) *Yesterday, here came the Princess, and as she passed hats were lifted.
(51) *Yesterday, there went Dr. and Mrs. Sorabjee, leaving little Amy alone at their table.

In DI, therefore, the verbal form indicates simultaneity with the assertion-time (be this coincident with the time of the utterance or not), a reading which is customarily expressed with progressive forms in English. Note, though, that if a progressive auxiliary were present in the Numeration, locality restrictions would prevent I-merge of the deictic predicate in Evid, since the auxiliary would be structurally closer to the probe than the main verb. This is why examples such as (52)–(55) below are not possible in English:

(52) *Here is coming the bus.
(53) *There is going Mary.
(54) *So I looked, and here was coming a white horse!
(55) *There were going Dr. and Mrs. Sorabjee, leaving little Amy alone at their table.

As expected, in LI, where the verb remains in the VP throughout the derivation (see 25), the impediment for progressive forms does not exist, something which De Wit (2016) has attested in an extensive corpus that she elicited from native speakers' surveys ((56) and (57) are her examples (33) and (34); also see sentence (24) above):

(56) In that house are living strange people.
(57) On top of the square block is lying another block.

The form and reading of the verb in DI are therefore crucially conditioned by the role of the predicate as the category that eventually encodes evidentiality in the derivation.

Another interesting prediction of the analysis in (30) has to do with the syntactic distribution of deictic inversion in English, which shows restrictions that are not present in its non-inverted counterpart or in LI. Once more, these restrictions can only be properly explained in terms of the specific pragmatic value of the construction.

DI, as all constructions involving full inversion in English (including LI), is a root/main clause phenomenon (Emonds 1970, 2004), and, as such, it occurs in main clauses, direct quotations, parentheticals and coordinate clauses:

(58) Here comes my bus.
(59) She said: "Here comes my bus".
(60) Here comes the bus, she said.
(61) I really should stay, but here comes my bus.

Nevertheless, while LI allows embedding in contexts which are root-like ("root-like indirect discourse embedding" contexts, RIDEs in Emond's terminology), DI heavily restricts embedding even in these cases:

(62) It seems that on the opposite corner stood a large Victorian mansion.
(63) *It seems that here comes my bus.

This again has to do with the fact that DI (contrary to LI) codifies illocutionary evidentiality, given that, as has been repeatedly claimed, illocutionary evidentials cannot embed (see Murray 2010; Demonte and Fernández-Soriano 2014, Aikhenvald 2015, and references therein). Significantly, the only context in which DI can be embedded and may still sound natural is the complement position of a perception verb, as in (64):[17]

(64) I can see that here comes my bus.

The subordination of DI to the verb *see* in (64) may be understood as the result of applying a double strategy of evidentiality (lexical and syntactic), through which the information source for the proposition is reinforced. The possibility to have DI as the complement of a lexical marker of visual evidentiality therefore provides additional support for the analysis of the construction as an evidential strategy.

It is also possible to find DI in peripheral adverbials which provide background propositions for the assertion in the main clause and are also root-like (see Haegeman 2004 for the distinction between central and peripheral adverbials in this respect); these clauses serve to structure the discourse, that is, to articulate the speech act, and are therefore compatible with an evidential strategy of this sort:

(65) I'd better leave, since here comes my bus.
(66) I'd stay a little longer, except here comes my bus.

It is interesting to note that Lakoff (1987, pp. 471–81), from where examples (61), (65) and (66) have been taken, aligns DI in terms of distribution with other constructions in English which convey assertions, such as negative questions (*Didn't Harry leave?*), inverted exclamations (*Boy! Is he ever tall!*), *wh*-exclamations (*What a fool he is!*), rhetorical questions (*Who on earth can stop Bernard?*) and reversal tags (*He is coming, isn't he?*). For him, the reason why all of these apparently unrelated constructions group together distributionally has to do with the fact that they are all speech act constructions, that is, constructions which are restricted in their use to expressing certain illocutionary forces. He claims that an adequate analysis of these constructions must necessarily pair their grammar with the illocutionary force they express, which is also the point I am making here.

In this regard, the impossibility to have DI in the interrogative or negative form also has to do with its illocutionary value as an evidential strategy. Whereas the non-inverted counterpart of the construction can be questioned (*Is the bus coming here?*), DI is used for the speaker to assert a proposition on the basis of some visual/sensory evidence; that is, the speaker is committed to the proposition, and this assertive value cannot be suspended. This is why the interrogative sentence in (67), though grammatical in English, cannot have an evidential reading (i.e., (67) is not a case of DI inversion, as the imperfective reading of the verb shows):

(67) Does the bus come here (regularly/* now)?

Additionally, illocutionary evidentiality does not contribute to the truth conditions of the proposition, and for the same reason, it cannot be accessed by linguistic operations bear-

ing on propositional truth, such as negation. DI, in particular, involves direct evidentiality, which implies that the evidential contribution (i.e., the fact that the speaker sees, hears, or feels something) can be neither challenged nor denied; in other words, the addressee cannot reply *no* or *that's not true* to mean that the speaker did not see/hear/feel that (see Murray 2021). Also note that EvidP, the category which finally hosts the verb, is structurally projected in the expressive component and thus outside the scope of the negative operator: negation can therefore access (some of the elements in) the proposition, but not the source of information for the proposition itself (see de Haan 1999, 2005; Demonte and Fernández-Soriano 2014; Murray 2021 among others).

The analysis of DI as a construction which marks evidentiality through syntactic means therefore formalizes its illocutionary force in discourse and accounts for its structural restrictions in a principled way. Obviously, this analysis should be further tested to fully confirm its empirical validity. Two questions immediately arise in this regard: is DI a syntactic strategy for evidentiality cross-linguistically, and (b) is DI the only construction where evidentiality is signalled syntactically? I offer a tentative answer for these questions here, leaving full treatment of the corresponding issues for future research.

With regard to the first (is DI a syntactic strategy for evidentiality cross-linguistically?), I expect DI to be possible in other non-evidential languages and have the same (or similar) restrictions that the construction manifests in English. At first sight, this seems to be the case for Spanish and probably other Indo-European languages as well.

Spanish word order is not as rigid as that of English and, as is well-known, the subject can be preverbal or postverbal in unmarked declarative sentences.[18] Postverbal subjects are also possible for discourse-dependent reasons, and the options here are also broader than in English. In the case of locative inversion, for example, not only unaccusatives (68) but also (in)transitive verbs (69) may undergo full inversion (see Ojea 2019 for details):

(68) En la puerta apareció una extraña criatura.
in the doorway appear-3SING.PAST a strange creature
'In the doorway appeared a strange creature'.

(69) En este garaje guarda Juan su bicicleta.
in this garage keep-3SING.PRES John his bicycle
'John keeps his bicycle in this garage'.

Spanish also displays a VS ordering in DI with *venir* and *ir*, a construction that has the same evidential reading as in English, that is, one in which the speaker brings the addressee's attention to an entity related to a proximal or distal location:

(70) Aquí viene el autobús.
here come-3SING.PRES thebus
'Here comes the bus'.

As in the case of English, the progressive forms that express ongoing situations in Spanish cannot be used in DI; as a matter of fact, verbal forms in DI cannot be analytic, even though full inversion with analytic forms is possible in other stylistic inversions, such as LI. Compare, in this respect, the DI examples (71) with standard cases of LI (72):

(71) a. *Aquí está viniendo el autobús.
here be-3SING.PRES come-PROGR the bus
'* Here is coming the bus'.
b. *Aquí ha venido el autobús
here have-3SING.PRES come-PERF the bus
'* Here has come the bus'.

(72) a. En mi jardín ya están floreciendo los rosales
in my garden already be-3SING.PL the rosebushes
flourish-PROGR
'Rosebushes are flourishing in my garden'.

b. En mi jardín ya han florecido los rosales
in my garden already have-3SING.PL the rosebushes
flourish-PERF
'Rosebushes have already flourished in my garden'.

Further, the tensed verbal form in Spanish DI must be understood as present/past of evidentiality with a reading of simultaneity with the assertion time, something we also observed for English. Therefore, the present form in the construction is not compatible with adverbs which express habitualness instead of simultaneity, such as *habitualmente* (*habitually*) (73); as expected, adverbs like this can freely modify LI structures (74):

(73) Aquí viene el autobús ahora /* habitualmente.
here come-3SING.PRES the bus now /usually
'Here comes the bus now/* usually'.

(74) En este terreno habitualmente florecen rosales
in this ground usually flourish-3PL.PRES rosebushes
'Rosebushes usually flourish in this ground'.

Similarly, past DI excludes adverbials which place the event as anterior to the time of the assertion, as is the case of *ayer* (*yesterday*) in (75):

(75) *Ayer ahí venía el bus.
yesterday there come-3SING.PAST.IMPERF the bus
'*Yesterday there came the bus'.

Note that only the imperfective past is possible in Spanish DI (as in (75)); the reason for this restriction is that Spanish perfective past focuses the limits of the event, and this makes it incompatible with the expression of simultaneity required by the evidential reading of the construction:

(76) Se fijó y, en efecto, ahí venía/* vino el bus.
look- and, in effect, there come- the bus
3SING.PAST.PERF 3SING.PAST.IMPERF/*come-
3SING.PAST.PERF
'She looked closely, and, in effect, there came the bus'.

And again, as expected, none of these restrictions are there in the structurally similar LI (i.e., nothing impedes past adverbials or the perfective past):

(77) En este terreno floreció el año pasado un rosal
in this ground flourish-3SING.PAST.PERFECT the year past a rosebush
'Last year a rosebush flourished in this ground'.

These facts therefore suggest that the derivation of DI in Spanish may also involve a discourse feature that drives the derivation and forces certain options over others. Hopefully, further investigation on DI in Spanish and other non-evidential languages may provide compelling evidence in favour of the status of the construction as a form of evidential marking.

As an anonymous reviewer has observed, it would also be interesting to explore how English DI is translated into proper evidential languages and check if the translation includes an evidential marker of some sort; if this were the case, the evidential status of DI would clearly be substantiated. Note, though, that the morphological marking of evidentiality is heavily language-dependent (i.e., there is not a systematic one-to-one correspondence between possible sources of evidence and morphology in evidential languages) and, as Aikhenvald (2015) mentions, evidential languages show fewer evidential distinctions in non-past tenses than in past tenses. Therefore, it could be the case that none of these evi-

dential languages has a specific morpheme to signal direct visual evidence in the present, but this will not necessarily constitute a counterargument to the existence of a syntactic strategy for it in other languages.

As for the second question (is DI the only construction where evidentiality is signalled syntactically?), one would expect evidentiality to play an active role in the syntax of some other constructions as well. Speas (2004, p. 258), following observations from Oswalt (1986) and Willett (1988), points out that the categories of evidentiality lie in a hierarchy which corresponds to the degree to which the evidence directly involves the speaker's own experience:

(78) Evidentiality hierarchy:
personal experience >> direct (e.g., sensory) evidence>> indirect evidence >> hearsay.

It is therefore important to explore not only if other constructions mark evidentiality through syntactic means in English, but also if syntactic evidentiality is subject to the same hierarchy found in the morphological system of evidential languages. As noted by an anonymous reviewer, if DI marks the kind of evidence which is at the top of the hierarchy (i.e., personal experience of the situation and direct evidence), the prediction will be that other constructions may mark the lower sources of evidence as well—i.e., indirect evidence.

Again, this seems to be the case. For example, Jiménez-Fernández and Tubino-Blanco (2023) offer a syntactic analysis of inferential questions in Spanish (whose main claims also apply to English: *What are you, on a diet?*) where indirect evidentiality plays an important role in the interpretation and form of these sentences. And, probably, word order in some of the constructions which Lakoff (1987, pp. 471–81) labels speech act constructions, such as negative questions (*Isn't it a beautiful day?*) or reversal tags (*He is coming, isn't he?*), could also be explained in terms of syntactic expression of indirect evidentiality.

4. Concluding Remarks

As Demonte and Fernández-Soriano (2014, p. 40) claim, investigation into the syntax–pragmatics interface must try to clarify the relevant elements which are explicitly elicited by the interaction between these components, and also the level of analysis which is most appropriate to characterize such interactions.

Relating to the first issue, the expression of the information source for the proposition through syntactic means constitutes a clear case of syntax–pragmatics interaction. One should note, though, that there exist different views on how to define evidentiality within the domain of grammar. For some scholars—mainly typologists—the notion of evidentiality should be restricted to the so-called evidential languages and, accordingly, to morphological marking; for others, evidentiality can be considered a more general functional category whose scope also includes lexical phenomena and can thus be extended to languages traditionally considered non-evidential. Squartini (2007), who offers an interesting account of these conflicting views, suggests that it is plausible to consider that morphological marking and lexical strategies might in fact be the opposite endpoints of a continuum which could admit intermediate stages, that is, linguistic forms less paradigmatic than evidential morphemes but more morphosyntactically constrained than, for example, evidential adverbs.

Adopting this intermediate view, languages which are not evidential in the strict sense may nonetheless employ syntactic and phonological means for the linguistic expression of evidentiality, and, as a result, the syntactic or phonological properties of some constructions in those languages may follow from the evidential reading that they have. I have suggested here that this is the case for deictic inversion, a construction whose word order and distribution can be explained in terms of a syntactic operation that places the verbal predicate high in the structure to signal direct evidentiality.

DI will then constitute a strategy of evidentiality in a language that does not codify evidentiality in the morphological system (i.e., which is not evidential in the strict sense). As I have shown, in this it differs from LI, which does not express the source of the information;

table (79) summarizes the main structural differences between the two constructions that follow from this fact:

(79)

	DI	**LI**
Type of verb	*Come*/*Go* Copula *be*	Unaccusative verbs of inherently directed motion, appearance or existence. Copula *be*
Initial locative constituent	*Here*/*There*	Any locative phrase
Expletive subject *there*	✗ Ex. (35)	✓ Ex. (26)–(28)
Analytic verbal forms	✗ Ex. (52)–(55)	✓ Ex. (24), (56), (57)
RIDE	✗ Ex. (63)	✓ Ex. (62)

The analysis of DI as an evidential strategy also accounts for its restricted distribution and for the otherwise unexpected difference between DI and its non-inverted counterpart in the reading of the present/past tense (cf., 43 and 48–51).

As regards the second issue (i.e., the level of analysis which is most appropriate to characterize syntax–pragmatics interactions), I have shown that the morphosyntactic expression of evidentiality in DI involves what Miyagawa (2022) terms the expressive component, that is, a structural layer "in the treetops" (i.e., above CP). A construction such as DI, whose grammar reflects its illocutionary force, therefore provides indirect evidence in favour of this upper level of structure which mediates between the act that the speaker engages in and the meaning of the utterance, encoding information about the speaker's commitment to the proposition and the information source for its content.

If this view is on the right track, the approach taken here to deictic inversion will hopefully pave the way for further research into the role of evidentiality and other discourse features in the syntactic derivation. Eventually, in-depth studies about the organization and structure of the expressive domain will serve to furnish principled explanations of a number of phenomena traditionally considered pragmatic and, consequently, to formalize the way in which central programmatic notions such as competence and performance interact.

Funding: This research has been funded by Spain's Ministry of Science, Innovation and Universities (MICINN), grant number PID2022-137233NB-I00.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are based on the author's own judgements or taken from publications cited and properly referenced.

Acknowledgments: I am indebted to the anonymous reviewers of *Language* for their very insightful remarks and suggestions, and to Julio Villa-García for fruitful discussion on the topic. I would also like to express my gratitude to the conference abstract reviewers and audience at the *IV Encontro de Gramática Gerativa* (Bahia, Brazil) and the *46th AEDEAN Conference* (Las Palmas de Gran Canaria, Spain). Needless to say, all remaining errors are exclusively my own.

Conflicts of Interest: The author declares no conflict of interest.

Notes

1 The syntactization of discourse is not unanimously accepted in the generative framework. See, for example, Fanselow (2008) and references therein for a defence of the hypothesis that syntax must be blind to categories of information structure.

2 Though eventually published in 2023, this work by Krifka has circulated extensively in its pre-print form since 2020.

3 SAP in Miyagawa's analysis actually stands for Speaker-Addressee Phrase, but it is equivalent in the relevant sense to the Speech Act Phrase in Speas and Tenny (2003).

4 Note, in this respect, that even if evidentiality, evaluative mood and epistemological modality are pragmatically connected, the existence of distinct heads for each of these notions has been shown to be necessary for certain grammatical accounts (for example, to provide a classification of logophoric predicates in terms of the projection they select; see Speas 2004).

5 For Murray (2010, p. 47), illocutionary evidentials are similar to certain English parentheticals (like . . ., *I hear*; . . ., *I find*; . . ., *they say*; . . ., *it's said*; . . ., *I take it*; . . ., *it seems*), while epistemic evidentials behave more like epistemic modals (such as *must*, *definitely*, *reportedly*, *apparently*. . .).

6 As Willett (1988, p. 51) points out, very few languages actually encode evidentiality as a separate grammatical category.

7 The abbreviations in examples (10) and (11), taken from Aikhenvald (2015, pp. 244–45), stand for SG (singular), REP (reported), O (object), CAUS (causative), IMMPST (immediate past) EYEWIT (eyewitness), M (masculine), DECL (declarative).

8 Examples taken from De Wit (2016, p. 110), Lakoff (1987, p. 504) and Kim (2003, p. 155).

9 As Lakoff (1987, pp. 469–70) points out, DI with copula *be* is very similar in its structure to the existential-*there* construction (e.g., *There is a man with a red hat on in the room*), to the point that it is even possible to find cases where both only differ in stress (stress indicated with capitals):

(i). THERE is an ape flirting with Harriet (DI)
(ii). There is an APE flirting with Harriet (Existential-*there* construction)

10 Examples taken from Ojea (2020).

11 For an alternative analysis where the locative phrase sits in TP at some point of the derivation, preventing the DP subject from moving there, see, among others, (den Dikken and Naess 1993; Bresnan 1994 and Rizzi and Shlonsky 2006).

12 Presentational sentences with a lexical or a covert expletive are not totally equivalent in structural terms, though: when the expletive is lexical (as in 26–28), it imposes a definiteness effect on its associate, which is not there if the expletive is covert. Therefore, a sentence such as (i) is grammatical in English, whereas (ii), its counterpart with the lexical expletive *there*, is not:
(i). In the top drawer of her desk lay her letter of resignation.
(ii). *In the top drawer of her desk there lay her letter of resignation.

13 Examples taken from an internet log: https://english.stackexchange.com/questions/246050/here-he-comes-here-comes-he-the-order-of-pronoun-and-verb-in-inversion (accessed on 10 November 2023).

14 One may object that in DI inversion with copula *be*, pronominal subjects cannot appear in the nominative:

(i). Here is me/* I and there is you

This is not construction-specific, though, but a general constraint on predicate nominals with copula *be* (*It is me/* I*; *The murderer is her/* she*); according to Newson (2018), this peculiarity has to do with the fact that the case system only sees arguments, and thus, predicate nominals of this sort get default case, which in English has the same form as the accusative.

15 This is not totally exceptional in English, and *come* and *go* pattern here with other cases of semantically light main verbs that can behave as auxiliaries and be subject to overt V to T (and T to C) movement, such as possessive *have* in some dialects (e.g., *Have you enough money?*).

16 I owe this observation to an anonymous reviewer of the paper.

17 I thank an anonymous reviewer, a native speaker of British English, for this observation. The same reviewer notes that a sentence such as (i) below could also be acceptable:

(i). ?She says that there goes her last chance at stardom.

Note, though, that the subordinate clause in (i) does not point at a real location shared by the speaker and the hearer (i.e., does not mark visual evidence), and therefore, its derivation could be different from that of standard perceptual deictic inversion. I leave the study of constructions of this type for future research.

18 See Fernández Ramírez (1986, p. 430 and ff.) for a very exhaustive description of the position of the subject in Spanish. As discussed there and in Ojea (2017) (and references therein), SV is not always the canonical position of the subject. As a matter of fact, VS is the default option when the subject is not the external argument of the verb; for example: (a) with psychological verbs such as *gustar* 'like', *preocupar* 'worry', or *molestar* 'bother', whose external argument is a dative experiencer (i); (b) with verbs such as *faltar* 'lack', *sobrar* 'excede', or *ocurrir* 'occur', whose external argument is a locative phrase which signals the place where the state or event originates (ii); and (c) with unaccusative verbs, the postverbal position being preferred in this case when the subject has an indefinite, set referring or existential reading (see Ojea 2017 for a description of the syntactic and semantic reasons which favour postverbal subjects of this type) (iii). The subject must also appear after the verb when there are

no postverbal modifiers, probably for prosodic reasons (e.g., the Weight to Prominence constraint; see Gutierrez-Bravo 2005) (iv):

(i).	Me	preocupa	tu	salud.			
	me	worry-3SING.PRES	your	health			
	'Your health worries me'.						
(ii).	Aquí	falta	un	tenedor.			
	here	miss-3SING.PRES	a	fork			
	'There is a fork missing here'.						
(iii).	A	esta	estación	llegan	solo	algunos	trenes.
	at	this	station	arrive-3PL.PRES	only	some	trains
	'Only some trains get to this station'.						
(iv).	Está	hablando	el	presidente.			
	be-3SING.PRES	talk-PROGR	the	president			
	'The president is talking'.						

As expected, the subject can also invert with the verb for reasons which have to do with information structure, as is the case of locative inversion or deictic inversion.

References

Aikhenvald, Alexandra Y. 2004. *Evidentiality*. Oxford: Oxford University Press.

Aikhenvald, Alexandra Y. 2015. Evidentials: Their links with other grammatical categories. *Linguistic Typology* 19: 239–77. [CrossRef]

Austin, John Langshaw. 1962. *How to Do Things with Words*. Oxford: Clarendon.

Birner, Betty J. 1996. *The Discourse Function of Inversion in English*. New York: Garland.

Boas, Franz. 1947. Kwakiutl Grammar, with a Glossary of the Suffixes. *Transactions of the American Philosophical Society* 37: 201–377. [CrossRef]

Boye, Kasper. 2012. *Epistemic Meaning. A Crosslinguistic and Functional-Cognitive Study*. Berlin and Boston: De Gruyter Mouton. [CrossRef]

Bresnan, Joan. 1994. Locative Inversion and the Architecture of Universal Grammar. *Language* 70: 72–131. [CrossRef]

Bruening, Benjamin. 2010. Language-Particular Syntactic Rules and Constraints: English Locative Inversion and Do-support. *Language* 86: 43–84. [CrossRef]

Chafe, Wallace, and Johanna Nichols, eds. 1986. *Evidentiality: The Linguistic Coding of Epistemology*. Norwood: Ablex. [CrossRef]

Chomsky, Noam. 2008. On Phases. In *Foundational Issues in Linguistic Theory: Essays in Honor of Jean-Roger Vergnaud*. Edited by Robert Robert Freidin, Carlos Peregrín-Otero and María Luisa Zubizarreta. Cambridge: MIT Press, pp. 133–66. [CrossRef]

Cinque, Guglielmo. 1999. *Adverbs and Functional Heads: A Cross-Linguistic Perspective*. Oxford: Oxford University Press. [CrossRef]

Coniglio, Marco, and Julia Zegrean. 2012. Splitting up force: Evidence from discourse particles. In *Main Clause Phenomena*. Edited by Lobke Aelbrecht, Liliane Haegeman and Rachel Nye. Amsterdam: John Benjamins, pp. 229–56. [CrossRef]

de Haan, Ferdinand. 1999. Evidentiality and Epistemic Modality: Setting the Boundaries. *Southwest Journal of Linguistics* 18: 83–101.

de Haan, Ferdinand. 2005. Encoding Speaker Perspective: Evidentials. In *Linguistic Diversity and Language Theories*. Edited by Zygmunt Frajzyngier, Adam Hodges and David S. Rood. Amsterdam: John Benjamins, pp. 379–97. [CrossRef]

de Vos, Mark. 2004. Pseudo coordination is not subordination. In *Linguistics in the Netherlands*. Edited by Leonie Cornips and Jenny Doetjes. Amsterdam: John Benjamins, pp. 181–92. [CrossRef]

De Wit, Astrid. 2016. The relation between aspect and inversion in English. *English Language and Linguistics* 20: 107–28. [CrossRef]

Demonte, Violeta, and Olga Fernández-Soriano. 2014. Evidentiality and Illocutionary force: Spanish Matrix *que* at the Syntax-Pragmatics Interface. In *Left Sentence Peripheries in Spanish. Diachronic, Variationist and Comparative Perspectives*. Edited by Andreas Dafter and Álvaro S. Octavio de Toledo y Huerta. Amsterdam: John Benjamins, pp. 217–52. [CrossRef]

Demonte, Violeta, and Olga Fernández-Soriano. 2022. A Multidimensional Analysis of the Spanish Reportative Epistemic Evidential *dizque*. *Lingua* 266: 103–68. [CrossRef]

den Dikken, Marcel, and Åshild Naess. 1993. Case dependencies: The case of predicate inversion. *Linguistic Review* 10: 303–36. [CrossRef]

Dorgeloh, Heidrun. 1997. *Inversion in Modern English. Form and Function*. Amsterdam: John Benjamins.

Emonds, Joseph. 1970. Root and Structure-Preserving Transformations. Ph.D. thesis, Massachusetts Institute of Technology, Cambridge, MA, USA.

Emonds, Joseph. 2004. Unspecified Categories as the Key to Root Constructions. In *Peripheries: Syntactic Edges and Their Effects*. Edited by David Adger and Cécile De Cat y George Tsoulas. Dordrecht: Kluwer, pp. 75–121. [CrossRef]

Faller, Martina. 2002. Semantics and Pragmatics of Evidentials in Cuzco Quechua. Ph.D. thesis, Stanford University, Stanford, CA, USA.

Faller, Martina. 2006. Evidentiality and Epistemic Modality at the Semantics/Pragmatics Interface. Paper presented at the Workshop on Philosophy and Linguistics, Ann Arbor, MI, USA, November 3–5.

Fanselow, Gisbert. 2008. In need of mediation: The relation between syntax and information structure. *Acta Linguistica Hungarica* 55: 397–413. [CrossRef]

Fernández Ramírez, Salvador. 1986. *Gramática española. Vol. 4: El verbo y la oración*. Madrid: Arco Libros.

González, Montserrat, Paolo Roseano, Joan Borràs-Comes, and Pilar Prieto. 2017. Epistemic and Evidential Marking in Discourse: Effects of register and debatability. *Lingua* 186–87: 68–87. [CrossRef]
Green, Georgia. 1982. Colloquial and Literary Uses of Inversion. In *Spoken and Written Language: Exploring Orality and Literacy*. Edited by Deborah Tannen. Norwood: Ablex, pp. 119–54. [CrossRef]
Gutierrez-Bravo, Rodrigo. 2005. Subject Inversion in Spanish Relative Clauses. A Case of Prosody-induced word Order Variation without Narrow Focus. In *Romance Languages and Linguistic Theory 2003*. Edited by Twan Geerts and Ivo van Ginneken y Haike Jacobs. Amsterdam: John Benjamins, pp. 115–28. [CrossRef]
Haegeman, Liliane. 2004. The Syntax of Adverbial Clauses and its Consequences for Topicalisation. *Antwerp Papers in Linguistics* 107: 61–90.
Haegeman, Liliane. 2014. West Flemish Verb-Based Discourse Markers and the Articulation of the Speech Act layer. *Studia Linguistica* 68: 116–39. [CrossRef]
Haegeman, Liliane, and Virginia Hill. 2013. The Syntactization of Discourse. In *Syntax and its Limits*. Edited by Raffaella Folli, Christina Sevdali and Robert Truswell. Oxford: Oxford University Press, pp. 370–90. [CrossRef]
Hoekstra, Teun, and René Mulder. 1990. Unergatives as Copular Verbs; Locational and Existential Predication. *Linguistic Review* 7: 1–79. [CrossRef]
Izvorski, Roumyana. 1997. The Present Perfect as an Epistemic Modal. In *Proceedings of SALT VIII*. Edited by Aaron Lawson. Ithaca: Cornell University Press, pp. 222–39. [CrossRef]
Jacobsen, William H., Jr. 1986. The Heterogeneity of Evidentials in Makah. In *Evidentiality. The Linguistic Coding of Epistemology*. Edited by Wallace Chafe and Johanna Nichols. Norwood: Ablex, pp. 3–28. [CrossRef]
Jakobson, Roman. 1957. *Shifters, Verbal Categories and the Russian Verb*. Cambridge: Harvard University Press.
Jiménez-Fernández, Ángel L., and Mercedes Tubino-Blanco. 2023. Inferential Interrogatives with qué in Spanish. *Languages* 8: 282–305. [CrossRef]
Jiménez-Fernández, Ángel L., and Shigeru Miyagawa. 2014. A feature-inheritance approach to root phenomena and parametric variation. *Lingua* 145: 275–302. [CrossRef]
Kaplan, David. 1989. Demonstratives. An Essay on the Semantics, Logic, Metaphysics and Epistemology of Demonstratives and other Indexicals. In *Themes from Kaplan*. Edited by Joseph Almog, John Perry and Howard Wettstein. Oxford: Oxford University Press, pp. 481–563.
Kay, Paul, and Laura Michaelis. 2017. Partial inversion in English. Paper presented at the 24th International Conference on Head-Driven Phrase Structure Grammar, Lexington, KY, USA, July 7–9; pp. 212–16. [CrossRef]
Kim, Jong-Bok. 2003. English Locative Inversion: A Constraint-Based Approach. *Korean Journal of Linguistics* 28: 207–35.
Krifka, Manfred. 2001. Quantifying into Question Acts. *Natural Language Semantics* 9: 1–40. [CrossRef]
Krifka, Manfred. 2015. Bias in Commitment Space Semantics: Declarative Questions, Negated Questions, and Question Tags. In *Semantics and Linguistic Theory (SALT) 25*. Phoenix: LSA Open Journal Systems, pp. 328–45. [CrossRef]
Krifka, Manfred. 2023. Layers of Assertive Clauses: Propositions, Judgements, Commitments, Acts. In *Propositionale Argumente Im Sprachvergleich: Theorie und Empirie/Propositional Arguments in Cross-Linguistic Research: Theoretical and Empirical Issues*. Edited by Jutta M. Hartmann and Angelika Wöllstein. Tübingen: Narr, pp. 116–83. [CrossRef]
Lakoff, George. 1987. *Women, Fire, and Dangerous Things. What Categories Reveal About the Mind*. Chicago: University of Chicago Press. [CrossRef]
Levin, Beth, and Malka Rappaport Hovav. 1995. *Unaccusativity at the Syntax-Lexical Semantics Interface*. Cambridge: MIT Press. [CrossRef]
Matthewson, Lisa, Henry Davis, and Hotze Rullmann. 2007. Evidentials as Epistemic Modals: Evidence from St'át'imcets. *Linguistic Variation Yearbook* 7: 201–54. [CrossRef]
Mélac, Eric. 2022. The Grammaticalization of Evidentiality in English. *English Language and Linguistics* 26: 331–59. [CrossRef]
Miyagawa, Shigeru. 2012. Agreements that Occur Mainly in Main Clauses. In *Main Clause Phenomena: New Horizons*. Edited by Lobke Aelbrecht, Liliane Haegeman and Rachel Nye. Amsterdam: John Benjamins, pp. 79–112.
Miyagawa, Shigeru. 2017. Linguistic Inquiry Monograph. In *Agreement Beyond Phi*. Cambridge: MIT Press, vol. 75.
Miyagawa, Shigeru. 2022. *Syntax in the Treetops*. Cambridge: MIT University Press.
Murray, Sarah E. 2010. Evidentiality and the Structure of Speech Acts. Ph.D. thesis, Rutgers University, New Brunswick, NJ, USA.
Murray, Sarah E. 2021. Evidentiality, Modality, and Speech Acts. *Annual Review of Linguistics* 7: 213–33. [CrossRef]
Newson, Mark. 2018. Default Case. *The Even Yearbook* 13: 29–55.
Nuyts, Jan. 2005. Modality: Overview and Linguistic Issues. In *The Expression of Modality*. Edited by William Frawley. Berlin and New York: Mouton De Gruyter, pp. 1–26. [CrossRef]
Ojea, Ana. 2017. Core intentional features in the syntactic computation: Deriving the position of the subject in Spanish. *Lingua* 195: 72–91. [CrossRef]
Ojea, Ana. 2019. EPP Satisfaction on Discourse Grounds: The Case of Locative Inversion. *Syntax* 22: 248–73. [CrossRef]
Ojea, Ana. 2020. Syntax Meets Discourse: Locative and Deictic (Directional) Inversion in English. *Poznań Studies in Contemporary Linguistics* 56: 251–76. [CrossRef]
Oswalt, Robert L. 1986. The evidential system of Kashaya. In *Evidentiality: The Linguistic Encoding of Epistemology*. Edited by Wallace Chafe and Johanna Nichols. Norwood: Ablex, pp. 20–29.

Palmer, Frank R. 1986. *Mood and Modality*. Cambridge: Cambridge University Press.
Portner, Paul, Miok Pak, and Raffaella Zanuttini. 2019. The Addressee at the Syntax-Semantics Interface: Evidence from Politeness and Speech Style. *Language* 95: 1–36. [CrossRef]
Postal, Paul. 1977. About a Nonargument for Raising. *Linguistic Inquiry* 8: 141–54.
Postal, Paul. 2004. A Paradox in English Syntax. In *Skeptical Linguistic Essays*. Oxford: Oxford University Press.
Reinhart, Tanya. 2006. *Interface Strategies: Optimal and Costly Computations*. Cambridge: MIT Press. [CrossRef]
Rizzi, Luigi. 1997. On the Fine Structure of the Left Periphery. In *Elements of Grammar*. Edited by Adriana Belletti and Luigi Rizzi. Dordrecht: Kluwer, pp. 281–337. [CrossRef]
Rizzi, Luigi, and Ur Shlonsky. 2006. Satisfying the Subject Criterion by a Non Subject: English Locative Inversion and Heavy NP Shift. In *Phases of Interpretation*. Edited by Mara Frascarelli. Berlin: De Gruyter, pp. 341–61. [CrossRef]
Rooryck, Johan. 2001. Evidentiality, Part I and Part II. *Glot International* 5: 125–33, 161–68.
Ross, John R. 1970. On declarative sentences. In *Readings in English Transformational Grammar*. Edited by Roderick A. Jacobs and Peter S. Rosenbaum. Waltham: Ginn, pp. 222–77. [CrossRef]
Sells, Peter. 1987. Aspects of Logophoricity. *Linguistic Inquiry* 18: 445–79. [CrossRef]
Sigurðsson, Halldór Ármann. 2004. The Syntax of Person, Tense and Speech Features. *Rivista di Linguistica* 16: 219–51.
Smith, Carlota S. 2000. *Accounting for Subjectivity (Point of View)*. Austin: Ms. University of Texas.
Speas, Margaret. 2004. Evidentiality, Logophoricity and the Syntactic Representation of Pragmatic Features. *Lingua* 114: 255–76. [CrossRef]
Speas, Peggy, and Carol Tenny. 2003. Configurational Properties of Point of View Roles. In *Asymmetry of Grammar*. Volume 1: Syntax and Semantics. Edited by Anna Maria DiSciullo. Amsterdam: John Benjamins, pp. 315–44. [CrossRef]
Squartini, Mario. 2007. Investigating a Grammatical Category and its Lexical Correlates. *Rivista di Linguistica* 19: 1–6.
Torrego, E. 1989. Unergative-unaccusative alternations in Spanish. In *Functional Heads and Clause Structure (MIT Working Papers in Linguistics 10)*. Edited by Itziar Laka and Anoop Mahajan. Cambridge: MIT Press, pp. 253–72.
Webelhuth, Gert. 2011. Motivating Non-Canonicality in Construction Grammar: The Case of Locative Inversion. *Cognitive Linguistics* 22: 81–105. [CrossRef]
Willett, Thomas. 1988. A Cross-Linguistic Survey of the Grammaticization of Evidentiality. *Studies in Language* 12: 51–97. [CrossRef]
Wiltschko, Martina. 2021. *The Grammar of Interactional Languages*. Cambridge: Cambridge University Press. [CrossRef]
Wiltschko, Martina, and Johannes Heim. 2016. The Syntax of Confirmationals. A Neo-performative Analysis. In *Outside the Clause. Form and Function of Extra-Clausal Constituents*. Edited by Gunther Kaltenböck, Evelien Keizer and Arne Lohmann Lohmann. Amsterdam: John Benjamins, pp. 305–40. [CrossRef]
Woods, Rebecca Louise. 2016. Investigating the Syntax of Speech Acts: Embedding Illocutionary Force. Ph.D. dissertation, University of York, York, UK.
Zanuttini, Raffaella. 2008. Encoding the Addressee in the Syntax: Evidence from English Imperative Subjects. *Natural Language & Linguistic Theory* 26: 185–218. [CrossRef]

languages

MDPI

Article

Middle-Passive Constructions, Dative Possessors, and Word Order in Spanish

Imanol Suárez-Palma

Department of Spanish and Portuguese Studies, University of Florida, Gainesville, FL 32611, USA; isuarezpalma@ufl.edu

Abstract: This paper examines data from Spanish middle-passive sentences whose grammatical subject contains a body-part noun, externally possessed by means of a dative possessor. I advocate for an analysis whereby the possessor originates inside the theme DP and raises to the specifier of an applicative projection to be licensed with dative case. I show that the unmarked order for dative DPs in these configurations is preverbal. These phrases may appear as the sole preverbal constituent, presumably in preverbal subject position, thus forcing the theme DP to remain inside the VP; alternatively, both the dative DP and theme DP can occur preverbally, in which case, the former appears to be left dislocated while the latter would be probed to preverbal subject position. This last scenario leads to a minimality violation, since the theme would be probed over the empty pronominal standing for the possessor that must necessarily sit in Spec, ApplP for the inalienable possession construal to obtain. Instead, I argue that both preverbal dative and theme DPs in Spanish middle-passive sentences are left dislocated and corefer with empty pronominals inside the sentence; the null dative possessor, being closer to T° always raises to subject position, which avoids any potential intervention effects. Finally, I explore how these data can be analyzed within a paratactic approach.

Keywords: dative possessors; middle-passive sentences; applicatives; minimality; clitic left dislocation

Citation: Suárez-Palma, Imanol. 2024. Middle-Passive Constructions, Dative Possessors, and Word Order in Spanish. *Languages* 9: 15. https://doi.org/10.3390/languages9010015

Academic Editor: Ana Ojea

Received: 22 October 2023
Revised: 21 December 2023
Accepted: 22 December 2023
Published: 27 December 2023

1. Introduction

The positions that preverbal subjects and dative DPs occupy in Spanish have been extensively discussed in the literature of generative grammar (Contreras 1976; Rivero 1980; Masullo 1992; Olarrea 1996; Fernández Soriano 1999; Ordóñez and Treviño 1999; Tubino 2007; Fernández Soriano and Mendikoetxea 2013; Villa-García 2015; Fábregas et al. 2017; Jiménez-Fernández 2020, to name a few); this paper aims to contribute to this debate by examining data from Spanish middle-passive sentences containing dative possessors DPs. Spanish middle-passive constructions are generic unaccusative predicates denoting intrinsic properties of the verb's internal argument, which must be a definite DP—never a bare NP—unmarkedly surfacing preverbally (1). These two properties are taken as evidence of this argument's externalization from the verbal domain (Suñer 1982; Fernández Soriano 1999), as well as its status as a sentential topic (Fodor and Sag 1982; Mendikoetxea 1999; Sánchez López 2002; Suárez-Palma 2019).

(1) *What happens?*
   a. Estos teléfonos se reparan fácilmente.
      these telephones RFL repair easily
      'These phones are easy to repair.'
   b. *Teléfonos se reparan fácilmente.
   c. ?Se reparan estos teléfonos fácilmente.[1]

This situation differs from what happens in other related unaccusative *se*-sentences, such as *se*-passives, where the grammatical subject commonly occurs postverbally and it can be a bare nominal (2).

(2) *What happens?*
a. Se repararon (estos) teléfonos.
RFL repaired these telephones
'(These) telephones were repaired.'
b. ?(Estos) teléfonos se repararon.

The verb's internal argument in middle passives may be a body-part noun. These are classified as relational nominals, for they inherently denote an inextricable part-whole or possession relationship with another entity; in other words, their meaning necessarily involves an inalienable possessor, unlike what happens with most common nouns (Picallo and Rigau 1999). In Spanish, the possessor may be expressed internally by means of a possessive determiner (3a) or a genitive PP (3b).

(3) a. Sus arrugas se ven fácilmente.
his wrinkles RFL see easily
'His wrinkles are easily visible.'
b. Las arrugas de Ismael se ven fácilmente.
the wrinkles of Ismael RFL see easily
'Ismael's wrinkles are easy to see.'

Alternatively, nouns favoring an inalienable possession reading can enter an external possession configuration, whereby the possessor is encoded as a dative argument of the verb in a sentential configuration. In Spanish, dative possessors are expressed by means of a dative clitic pronoun, which can be optionally doubled by a dative DP (Kliffer 1983; Demonte 1988; Kempchinsky 1992; Gutiérrez Ordóñez 1999; Picallo and Rigau 1999; Guéron 2006; Sánchez López 2007; Conti 2011). In active contexts, the dative possessor DP usually appears after the verb in out-of-the blue contexts, as shown in (4).

(4) *What happens?*
a. Minerva le_i vio [las arrugas]$_i$ **a Albus$_i$**.
Minerva 3SG.DAT saw the wrinkles Albus.DAT
'Minerva saw Albus' wrinkles.'
b. Minerva le_i vio **a Albus$_i$** [las arrugas]$_i$.
c. ?**A Albus$_i$** Minerva le_i vio [las arrugas]$_i$.
d. ?**A Albus$_i$** le_i vio [las arrugas]$_i$ Minerva.

On the contrary, the unmarked position for the dative DP in out-of-the-blue middle-passive contexts is preverbal, either as the sole fronted constituent, therefore forcing the theme DP to remain inside the VP (5a), or together with the theme DP (5b,c), but not postverbally (5d).

(5) *What happens?*
a. **A Ismael$_i$** se le_i ven [las arrugas]$_i$ fácilmente.
Ismael.DAT RFL 3SG.DAT see the wrinkles easily
'Ismael's wrinkles are easy to see.'
b. **A Ismael$_i$**, [las arrugas]$_i$ se le_i ven fácilmente.
c. [Las arrugas]$_i$, **a Ismael$_i$** se le_i ven fácilmente.
d. ?[Las arrugas]$_i$, se le_i ven **a Ismael$_i$** fácilmente.

The idiosyncrasy of middle-passive sentences with respect to their word order, along with their interaction with dative arguments, can therefore be used to gain further insight into the hotly debated positions that preverbal subjects and dative DPs occupy in Spanish.

Among the different generative analyses of dative possessors in Romance languages in general, and in Spanish in particular, Cuervo (2003) proposes that these DPs are introduced in the specifier of a low applicative projection (Pylkkänen 2002), i.e., an argument-introducing functional head responsible for relating two entities: a possessor in its specifier and a possessee in its complement position. The dative clitic pronoun, whose phi features match those of the possessor DP, spells out the applicative head, and the entire applicative projection merges as the verb's complement. Because middle-passive configurations lack

an external argument, T° would probe the closest DP—the dative DP—to its specifier, while the theme DP would remain inside the VP, as shown in (6).

(6) [$_{TP}$ **a Ismael$_i$** [$_T$ se le$_i$ ven [$_{VoiceP}$ [$_{\upsilon P}$ [$_{\sqrt{P}}$ [$_{ApplP}$ ***t$_i$*** [$_{Appl}$ ~~le~~ [$_{DP}$ las arrugas]]] $\sqrt{}$~~-~~]]]]]

While this approach accounts for the sentence in (5a) straightforwardly, I will show that it runs into a minimality violation (Rizzi 1990, 2012) when deriving the sentence in (5b), where it appears that the dative DP is left dislocated, and the theme DP sits in the preverbal subject position. If that was the case, the theme DP would be probed to subject position over the empty dative pronoun that must necessarily sit in the specifier of the applicative head for the relationship of possession to hold.

(7) **a Ismael$_i$** [$_{TP}$ las arrugas$_k$ [$_T$ se le$_i$ ven [$_{VoiceP}$ [$_{\upsilon P}$ [$_{\sqrt{P}}$ [$_{ApplP}$ ***pro$_i$*** [$_{Appl}$ ~~le~~ [$_{DP}$ ***t$_k$***]]] $\sqrt{}$~~-~~]]]]]

Instead, I will present data suggesting that preverbal dative DPs and theme DPs in Spanish middle passives are left dislocated. This will lead me to pursue an analysis within the minimalist generative framework (Chomsky 1995) that is in line with Barbosa's (2009) account for preverbal subjects in consistent null subject languages like European Portuguese, whereby these constituents are left dislocations coreferring with empty pronouns inside the sentence. I will show how this proposal avoids any potential intervention effects. Finally, I will also explore how these data can be accommodated within a biclausal/paratactic approach (Ott 2014, 2015; Fernández-Sánchez and Ott 2020; Villa-García and Ott 2022, *inter alia*).

The article is structured as follows: Section 2 discusses how inalienable possession can be analyzed in Spanish, focusing particularly on instances of external possession. Section 3 is devoted to Spanish middle-passive sentences; in it, I provide a brief survey of the most salient structural properties of middle-passive sentences in Spanish and show how they can be analyzed syntactically. In Section 4, I describe how middle-passive sentences interact with dative possessors, paying special attention to the different possible word orders; moreover, I explain why a low applicative analysis of dative possessors in these constructions along the lines of Cuervo (2003) is susceptible to run into intervention effects. I contend that this technical shortcoming can be done away with if preverbal dative and theme DPs are extrasentential, i.e., left dislocations, and I provide evidence to support this idea. Finally, Section 5 concludes the paper.

2. Dative Possessors and (in-)Alienable Possession

The inextricable connection between a body-part noun and its possessor, or that between two entities in a part-whole relation, is known as inalienable possession (Guéron 2006). This construal is also possible with relational nouns, i.e., nominals denoting items pertaining to someone's personal sphere (Bally 1926), including those referring to personality traits, family members, and familiar objects such as items of clothing. The grammar of Spanish offers different strategies to encode (in)alienable possession. On the one hand, the possessor—be it inalienable or not—can be conveyed internally, i.e., inside the DP containing the possessum, by means of a possessive determiner (8a), a strong possessive (8b),[2] or as a DP inside a PP (8c).

(8) a. Sus arrugas/cartas.
his.PL wrinkles/letters
'His wrinkles/letters.'

b. Las arrugas/cartas suyas.
the.F.PL wrinkles/letters his.F.PL
'His wrinkles/letters.'

c. Las arrugas/cartas de Javier.
the wrinkles/letters of Javier
'Javier's wrinkles/letters.'

On the other hand, inalienable possession in Spanish and other Romance languages is frequently associated with external possession (Kliffer 1983; Demonte 1988; Kempchinsky 1992; Gutiérrez Ordóñez 1999; Picallo and Rigau 1999; Sánchez López 2007; Conti 2011),[3] i.e., a sentential configuration in which both the possessor and the possessum surface independently as two different arguments of the verb, the former marked with dative case, while the latter bears accusative case (9). In most Spanish dialects, the dative possessor generally appears to be in complementary distribution with possessive determiners heading the possessum in these contexts;[4] instead, the latter is commonly headed by a definite determiner or, alternatively, a quantifier (Demonte 1988; Kempchinsky 1992; Picallo and Rigau 1999; Gutiérrez Ordóñez 1999). The dative argument is therefore understood as the possessor or the location of the possessum (Cuervo 2003).

(9) Tania le_i vio [*sus_i/las/varias arrugas]$_i$ (a $Alberto_i$).
Tania 3SG.DAT saw his/the/several wrinkles.ACC Alberto.DAT
'Tania saw (several of) Alberto's wrinkles.'

External possession in Romance languages has been analyzed in terms of binding and control (Guéron 1983, 1985; Demonte 1988), predication (Vergnaud and Zubizarreta 1992), possessor raising (Demonte 1995; Sánchez López 2007; Nakamoto 2010; Suárez-Palma 2022; Rodrigues 2010, 2023), applied arguments (Cuervo 2003), or a mixture of the last two approaches, i.e., possessor raising to the specifier of an applicative projection (Armstrong 2021; Suárez-Palma forthcoming).

In her theory of dative arguments in Spanish, Cuervo (2003) establishes a parallelism between dative possessors and datives in double object constructions (henceforth, DOCs), and notes that these arguments are structurally identical in terms of case, hierarchical position, word order, and spell-out form, i.e., the dative clitic. Moreover, dative possessors and datives in DOCs are semantically related directly to the object, not the verb. Thus, Cuervo concludes that Spanish dative possessors are indeed instances of DOCs.

(10) Tania le_i envió un mensaje a $Luis_i$.
Tania 3SG.DAT sent a message.ACC Luis.DAT
'Tania sent Luis a message.'

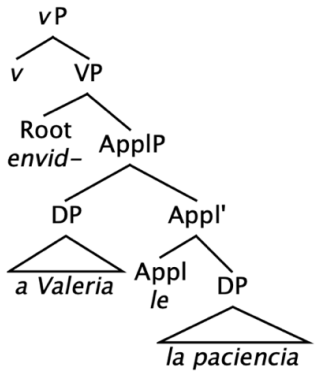
Cuervo assumes Pylkkänen's (2002) analysis of DOCs, whereby dative arguments in these configurations are introduced in the specifier of a low applicative head, i.e., an argument-introducing functional head relating two arguments, the dative DP in its specifier, and the theme in its complement position; the entire ApplP merges as the verb's internal argument. According to Pylkkänen, the particular semantics of the applicative determines whether the argument in its specifier is interpreted as a goal (11a) or a source (11b). Furthermore, Cuervo adds a third kind of low applicative head, a possessor applicative, whose semantics convey a static relation of possession, and which is responsible for introducing dative possessors (11c).

(11) a. $APPL_{TO}$ (Goal applicative):
$\lambda x.\lambda y.\lambda f_{<e<s,t>>}.\lambda e.$ f(e,x) & theme (e,x) & from-the-possession-(x,y)
b. $APPL_{FROM}$ (Source applicative):
$\lambda x.\lambda y.\lambda f_{<e<s,t>>}.\lambda e.$ f(e,x) & theme (e,x) & to-the-possession-(x,y)
c. $APPL_{AT}$ (Possessor applicative):
$\lambda x.\lambda y.\lambda f_{<e<s,t>>}.\lambda e.$ f(e,x) & theme (e,x) & in-the-possession-(x,y)

Thus, Cuervo analyzes sentences containing a possessor dative as shown in (12).

(12) a. Cuervo (2003, p. 76; example 86a)
Pablo le_i envidia [la paciencia]$_i$ a $Valeria_i$.
Pablo 3SG.DAT envies the patience Valeria.DAT
'Pablo envies Valeria for her patience.'

b.



In (12), the root $\sqrt{}$*envid-* takes a low applicative of possession (Low-APPL$_{AT}$) projection, in whose specifier merges the dative DP *a Valeria* that is related to the DP *la paciencia* in complement position, thus allowing the relationship of possession to obtain; the dative clitic *le*, whose phi features mirror those of the dative DP, spells out the applicative head.

Cuervo points out that the dative possessor construction is not restricted to inalienable possession and proposes the same derivation for sentences like (13), where the possessum is alienably possessed. In her proposal, the applicative is needed to establish the possessive relationship between the possessor originating in its specifier and the possessee in its complement position. Cuervo does not delve into how the inalienable possession construal arises in these structures, and one must assume it is determined by the type of noun the possessum DP contains.

(13) Cuervo (2003, p. 74; example 87a)

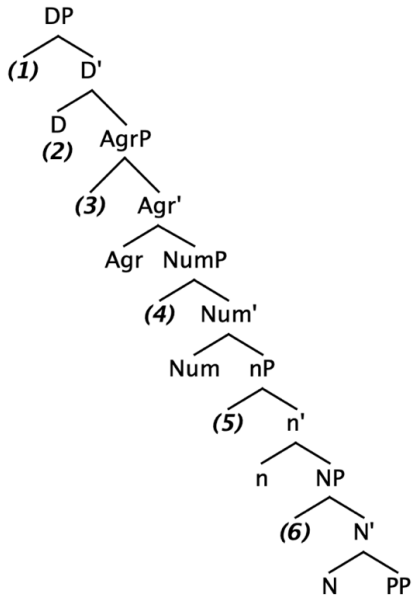
Pablo le$_i$	envidia	[el	auto]$_i$	a Valeria$_i$.	
Pablo 3SG.DAT	envies	the	car	Valeria.DAT	

'Pablo envies Valeria for her car.'

Here, I will adopt Suárez-Palma's (forthcoming) analysis of possession, which is based on an adaptation of Alexiadou et al.'s (2007) proposal of the possible positions where possessors can merge or be licensed inside the possessum DP, captured in (14).

(14) Adapted from Alexiadou et al. (2007, p. 575)

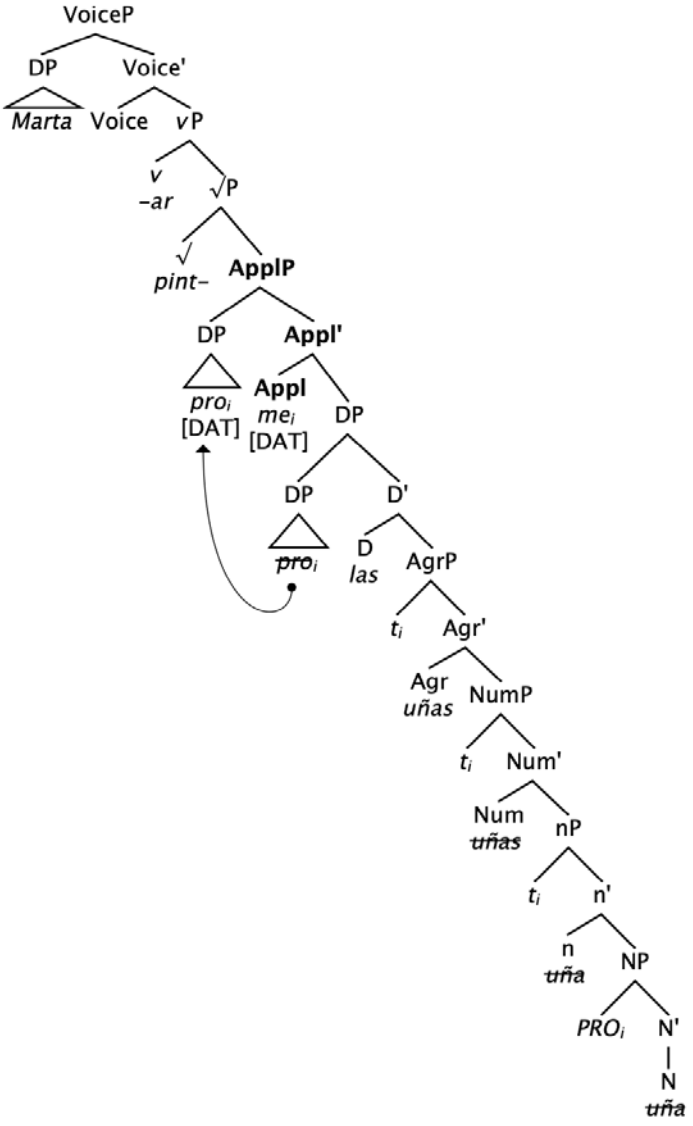
1.	Lexical DP possessives	*John's book*	(English)
2.	Clitic possessives	*Su libro*	(Spanish)
3.	'Weak' pronoun possessives	*El seu llibre*	(Catalan)
4.	Postnominal strong possessors	*El seu llibre*	(Spanish)
5.	Alienable possessors		
6.	Inalienable possessors (PRO)		



Suárez-Palma argues that all possessors originate in the specifier of Spec,nP, and later move within the DP to be licensed with genitive case. What distinguishes alienable from inalienable possession is the fact that relational and body-part nouns take a PRO as an argument in their specifier, which is controlled by the possessor in Spec,nP.[5] The latter can be an empty pronominal *pro* or a full DP, both of which will require case licensing. When genitive case is available inside the possessum DP, *pro* will surface as a clitic (*sus ojos*; 'her eyes'), weak (*els seus ulls*; 'her eyes'),[6] or strong possessive (*los ojos suyos*; 'her eyes'), depending on the position this argument raises to within the DP. On the other hand, if it is a full DP, it will be case-marked by genitive preposition *de* (*los ojos de Marina*; 'Marina's eyes').[7] In the event that genitive case is not available inside the possessum DP, the possessor DP will need to exit it and raise to a position where it can be case-licensed, i.e., the specifier of an applicative phrase. In other words, for Suárez-Palma, the function of the low applicative of possession is merely to case-license the possessor DP when no functional projection inside the possessum DP is able to do so. This straightforwardly accounts for the complementary distribution between possessives and dative possessors in most Spanish dialects and in other Romance languages.[8] Thus, a sentence containing a dative possessor would show the derivation in (15).

(15) a. Marta me_i pintó [las uñas]$_i$.
Marta 1SG.DAT painted the nails
'Marta painted my nails.'

b.



In (15b), the relational noun *uña* takes PRO as an argument in its specifier, from where it is controlled by the empty pronominal *pro*, originating in Spec,nP, and standing for the possessor. The noun undergoes head movement to Agr° from where it establishes an agreement relation with the c-commanding determiner; *pro*, on the other hand, raises to the specifier of the low applicative head—spelled out by the dative clitic pronoun *me*—to value dative case. The dative clitic incorporates into the verb, which raises to T°; finally, the external argument *Marta* is introduced in the specifier of a Voice projection and is later probed by T° to its specifier to check its EPP feature, where it is also assigned nominative case. In the next section, I describe the most salient properties of middle-passive sentences in Spanish and discuss how these configurations interact with dative possessors.

3. Middle-Passive Constructions in Spanish: Description and Analysis

Cross-linguistically, the middle voice refers to a number of stative, generic configurations denoting atemporal intrinsic properties of the verb's internal argument, which surfaces as the grammatical subject (Ackema and Schoorlemmer 2006, inter alia).

(16) Cotton shirts iron easily.

Languages differ in the way their grammars encode this construal. Lekakou (2005), for instance, explains that middle constructions in languages such as Dutch, German, or English are syntactically unergative, while in others like Greek or French, they are unaccusative predicates, being syntactically indistinguishable from generic passives. Such crosslinguistic variation resulted in the development of numerous analyses of different natures, including syntactic (Keyser and Roeper 1984; Hale and Keyser 1988; Roberts 1987; Stroik 1992; Lekakou 2005; Schäfer 2008; Suárez-Palma 2019; Suárez-Palma 2020; Fábregas 2021), semantic (O'Grady 1980; Dixon 1982; Condoravdi 1989; Chierchia 2003; Lekakou 2005), and lexicalist accounts (Fagan 1992; Ackema and Schoorlemmer 1995), to name a few. Despite this heterogeneity, there is consensus in the literature regarding several common traits these structures share across languages, namely the lack of an explicit external argument, the internal argument's promotion to grammatical subject, their generic, nonepisodic nature, their modal interpretation, as well as the quasi-mandatory modification by an adjunct.

The structures under consideration here have been traditionally known as middle-passives in the canonical descriptive work on Spanish grammar (Mendikoetxea 1999). These sentences contain the reflexive clitic pronoun *se* in its third person form, and their generic, stative nature makes them compatible only with imperfective tenses, i.e., present or imperfect. Moreover, middle-passive constructions denote the participation of a generic implicit agent in the event, which can be rephrased as 'anyone,' although its explicit realization by means of a by-phrase is banned.

(17) Esta blusa se lava fácilmente (*por Pedro).
this blouse RFL washes easily by Pedro
'This blouse washes easily.'
'Anyone can wash this blouse easily.'

Mendikoetxea (1999) argues that the implicit external argument in middle-passives must necessarily be an agent; therefore, only verbs denoting activities or achievements would be eligible to enter these configurations (18a), while those whose external arguments are experiencers would be ungrammatical, as shown in (18b), which contains a predicate denoting a durative accomplishment (18b).

(18) a. La historia de España se aprende fácilmente.
the history of Spain RFL learns easily
'The history of Spain is easy to learn.'
b. Mendikoetxea (1999, p. 1656)
*La historia de España se Sabe de memoria.
the history of Spain RFL knows of memory
Intended: 'The history of Spain is known by heart.'

The lack of an explicit external argument in middle-passive configurations favors the promotion of the verb's internal argument to grammatical subject, triggering agreement with the verb; in fact, Hale and Keyser (1988) consider this argument to be a semiagent in these sentences, since its intrinsic properties are somehow responsible for the event. In (18a), for instance, it is the idiosyncratic properties of the history of Spain that favor its learnability.

As mentioned above, the grammatical subject in middle passives occurs preverbally in unmarked contexts and cannot be a bare NP, as shown by the ungrammaticality of (19a) should the determiner *estas* be removed; this has been interpreted as evidence for the internal argument's externalization from the VP (Suñer 1982; Fernández Soriano 1999).[9] The reason for this externalization would be the fact that the grammatical subject in middle-passive sentences has a discursive function, i.e., it is a sentential topic (Fodor and Sag 1982; Mendikoetxea 1999; Sánchez López 2002; Suárez-Palma 2019).

(19) *What happens?*

a. Estas blusas se lavan fácilmente.
these blouses RFL wash easily
'These blouses wash easily.'

b. *Blusas se lavan fácilmente.

c. ?Se lavan estas blusas fácilmente.[10]

However, Mendikoetxea (1999) notes that this argument remains inside the verbal domain when another constituent is focalized and fronted (20).

(20) EN LA LAVADORA se lavan **estas blusas** fácilmente, no a mano.
in the washer RFL wash these blouses easily not by hand
'IN THE WASHER these blouses wash easily, not by hand.'

In this respect, middle passives differ from other unaccusative *se*-sentences like *se*-passives, whose grammatical subject—which is also the verb's internal argument—can be a bare NP and tends to occur postverbally (21).

(21) *What happens?*

a. Se lavaron blusas.
RFL washed blouses
'Blouses were washed.'

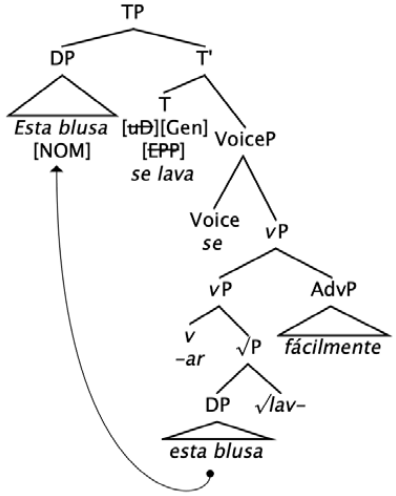
b. *Blusas se lavaron.

Finally, it is generally agreed that middle-passive constructions convey a modal interpretation (Mendikoetxea 1999; Sánchez López 2002), which is evidenced by the fact that they can be rephrased as a prototypical modal sentence (e.g., *anyone can wash these blouses easily*). Moreover, these structures are often modified by an adverbial or prepositional phrase, which enhances their modal reading.[11]

Considering all the above, a middle-passive sentence like the one in (22) would show the following derivation:

(22) a. Esta blusa se lava fácilmente.

b.



In (22), I follow Schäfer (2008) and Suárez-Palma (2019, 2020) in assuming that middle-passive constructions contain a Voice head (Kratzer 1996), which is spelled out by the reflexive clitic *se* and does not introduce an external argument. Furthermore, the combination of a generic operator (Gen) and an uninterpretable [D] feature in T° would cause the probing of the only DP available in the derivation, i.e., the theme DP *esta blusa*, to the specifier of the TP to cancel its EPP feature, thus becoming the grammatical subject and valuing nominative case. Finally, the root $\sqrt{}$*lav-* undergoes head movement, incorporating any clitics it finds on its way to T°.

In this section, I have outlined some of the most salient structural properties of middle-passive sentences in Spanish. For a more in-depth examination of the latter, I refer the reader to the thorough descriptions by Mendikoetxea (1999) and Sánchez López (2002), or the work by Suárez-Palma (2019, 2020) and Fábregas (2021) for more recent discussions. Next, I will discuss the interaction between middle-passive sentences and dative possessors in Spanish, paying particular attention to issues concerning word order in these configurations.

4. Dative Possessors, Middle-Passive Constructions and Word Order

The position of preverbal dative DPs in Spanish has been a highly debated topic in the literature (Masullo 1992; Fernández Soriano 1999; Tubino 2007; Fernández Soriano and Mendikoetxea 2013; Fábregas et al. 2017, among others). For instance, in his study of preverbal and postverbal datives, Jiménez-Fernández (2020) argues that dative possessors always surface postverbally in out-of-the-blue contexts (23), because they are generated low in the structure by means of a low applicative head, as proposed by Cuervo (2003).

(23) Jiménez-Fernández (2020, p. 240; ex. 71)
What's up?
a. Le$_i$ besé [la mano]$_i$ a María$_i$.
3SG.DAT kissed the hand María.DAT
'I kissed María's hand.'
b. ?A María le besé la mano.

Nevertheless, this does not seem to be the case for dative possessors in middle-passive configurations, which unmarkedly occur preverbally, presumably due to the lack of an external argument in these structures.

(24) *What happens?*
a. A Ismael$_i$ se le$_i$ ven [las arrugas]$_i$ fácilmente.
Ismael.DAT RFL 3SG.DAT see the wrinkles easily
'Ismael's wrinkles are easy to see.'
b. A Ismael$_i$, [las arrugas]$_i$ se le$_i$ ven fácilmente.
c. [Las arrugas]$_i$, a Ismael$_i$ se le$_i$ ven fácilmente.
d. ?[Las arrugas]$_i$, se le$_i$ ven a Ismael$_i$ fácilmente.

In this respect, dative possessors in Spanish middle-passive constructions resemble preverbal dative experiencers. Masullo (1992) notes that negatively quantified dative experiencer DPs, which also tend to occur preverbally, lose their quantificational scope if they are left dislocated, and are thus interpreted referentially, as shown in (25).

(25) Masullo (1992, p. 90)
a. A nadie le gusta la música pop en esta casa.
nobody.DAT 3SG.DAT likes the music pop in this house
'Nobody likes pop music in this house.'
b. *A nadie, le gusta la música pop en esta casa.
'*Nadie* likes pop music in this house.'

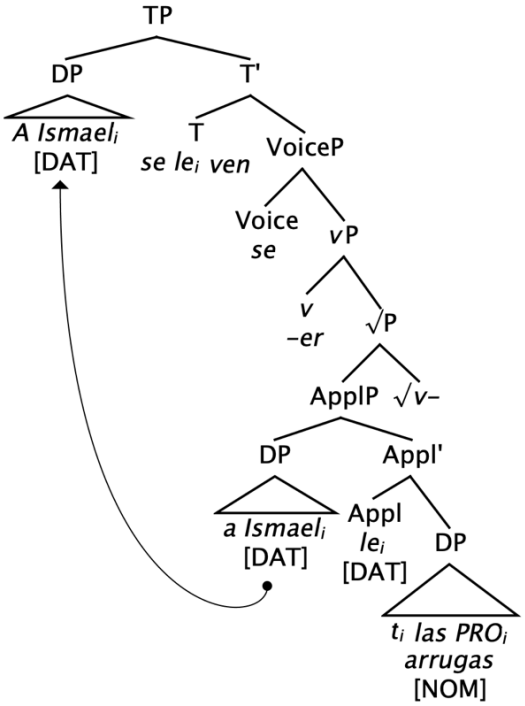
In (25a), the dative DP must be sitting in an A-position inside the sentence, hence its quantificational scope obtains. However, in (25b), the dative phrase is left dislocated outside the TP, therefore losing such interpretation. Masullo takes this as evidence for the fact that these phrases have subject-like properties and occupy the preverbal subject

position, presumably Spec,TP. Interestingly, the same phenomenon is attested with dative possessor DPs in Spanish middle-passive sentences.

(26) a. A nadie$_i$ se le$_i$ ven [las arrugas]$_i$ fácilmente.
nobody.DAT RFL 3SG.DAT see the wrinkles easily
'Nobody's wrinkles are easily visible.'
b. *A nadie, las arrugas se le ven fácilmente.
'*Nadie*'s wrinkles are easily visible.'

The dative DP in (26a) appears to have raised to preverbal subject position, where the quantificational reading of *nadie* ('nobody') obtains. The theme DP is forced to remain inside the VP, which is not normally the case in Spanish middle-passive sentences without dative DPs. On the other hand, the dative DP in (26b) is left dislocated, which results in the loss of its quantificational reading, and allows the theme's promotion to a preverbal subject position.[12] Note that despite the ungrammaticality of (26b), the dative DP and the theme may both surface together before the verb when no quantifiers are involved, as shown in (24b) and (24c) above. Considering the data above, a plausible derivation of a middle-passive sentence containing a preverbal dative possessor DP would be the following:

(27) a. A Ismael$_i$ se le$_i$ ven [las arrugas]$_i$ fácilmente.
Ismael.DAT RFL 3SG.DAT see the wrinkles easily
'Ismael's wrinkles are easy to see.'
b.

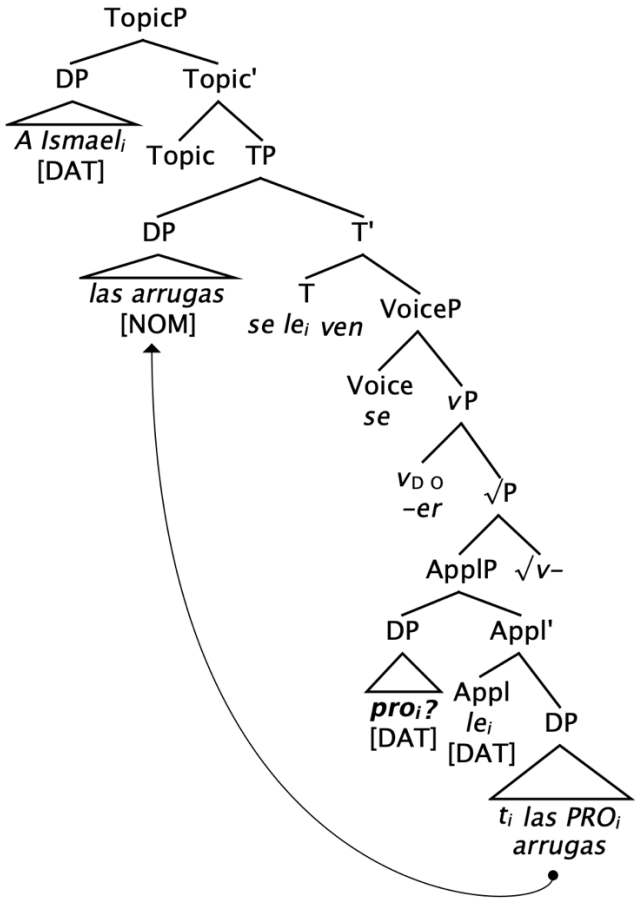


In (27), the possessor DP *Ismael* originates inside the possessum DP, in Spec,nP, from where it controls the PRO in the specifier of the relational noun *arrugas*. The possessor DP later exits the possessum DP to raise to Spec,ApplP, where it values dative case; the entire applicative projection merges as the complement of the verbal root $\sqrt{v}$-. Because this is a middle-passive configuration, no external argument is projected in the specifier of VoiceP, which is spelled out by *se*. In order to value its EPP feature, T° probes the closest DP to its specifier: in this case, the dative DP *a Ismael*; because this DP is already case-marked, T° assigns nominative case to the theme DP *las arrugas* via Agree. Finally, the verb undergoes head movement to T, incorporating all the clitics it finds on its way.

While the derivation in (27b) straightforwardly accounts for middle-passive sentences where the possessor dative DP is the sole preverbal constituent, a problem arises in cases where both the dative DP and theme DP occur preverbally, as is shown in (28).

(28) a. A Ismael$_i$, [las arrugas]$_i$ se le$_i$ ven fácilmente.

b.



In (28), the dative DP *a Ismael* merges outside the TP as a dislocated constituent in the left periphery (Rizzi 1997), here in Spec,TopicP, and bears an identity relation with the dative clitic *le* inside the sentence; in other words, this is a clitic left-dislocation (CLLD) configuration (Cinque 1990). The verbal root $\sqrt{v}$- takes a low applicative head as its complement, containing the possessee *las arrugas*. Because the dative DP is dislocated, the clause-internal argument standing for the possessor inside the sentence must be a null pronoun *pro* which is unable to license genitive case inside the possessum DP, and therefore raises to the specifier of the low applicative phrase. Note that assuming that nothing sits in Spec,ApplP would violate the semantic definition of the low applicative of possession (low-APPL-AT) given in (11c). If *pro* sits in Spec,ApplP then, when T° looks down to probe a DP to its specifier, it would find *pro* rather than the theme, since the former is the closest DP to T°. In other words, probing the possessee over *pro* would lead to a violation of minimality (Rizzi 1990). In the next section, I explore a possible way to avoid these intervention effects.

4.1. Intervention Effects: A Solution

As seen above, a low applicative analysis of dative possessors in middle-passive configurations is likely to run into a locality violation when dealing with the configuration in which both the dative DP and theme DP occur preverbally, if we assume that the former is left dislocated and the latter sits in Spec,TP. An empty pronoun standing for the possessor must sit in Spec,ApplP in order to abide by the semantic notation of the low applicative of possession (low-APPL-AT) (cf. (11c) above), and this argument would intervene when T° tries to probe the theme DP sitting in Appl°'s complement position. In order to overcome this technical obstacle, I propose that dative DPs and preverbal subjects in middle-passive constructions are left-dislocated constituents, base-generated outside of the sentence, and coreferencing with empty pronominals in argument positions; in other words, these are

instances of CLLDs. Rigau (1988) noted the different distributions that *pro* and lexical subjects have, and showed that *pro*'s behavior parallels that of clitics, not strong pronouns. Similarly, Olarrea (2012) explains that the coreferential element in CLLDs has to be an empty pronominal licensed by agreement, or a clitic, but never a tonic pronoun or a phrase (29).

(29) *Con Chloe$_i$, siempre viajo con Chloe$_i$/ella$_i$.
with Chloe always travel with Chloe/she
'I always travel with Chloe/her.'

Finally, Baker (1995) proposed that lexical DP arguments are always associated with a *pro*, which is the real argument, while lexical DPs are adjoined to a more peripheral position; he argued that the latter are not derived by movement but computed representationally through coindexation, following Cinque's (1990) intuition for CLLDs. Let us look at the data more closely to see whether this proposal is on the right track.

(30) a. [Las arrugas]$_i$, a nadie$_i$ se le$_i$ ven fácilmente.
the wrinkles nobody.DAT RFL 3SG.DAT see easily
'Nobody's wrinkles are easy to see.'
b. **Las arrugas$_k$** [$_{TP}$ a nadie$_i$ se le$_i$ ven ***pro*$_k$** fácilmente]

In (30), the dative DP *a nadie* must be sitting in the preverbal subject position, i.e., Spec,TP, since a quantificational reading obtains; this implies that the theme DP *las arrugas* is therefore left dislocated. If we assume a base generation approach for left-dislocated constituents (Cinque 1990; Frascarelli 1997, 2000),[13] then a third person plural null object pronoun *pro* coreferring with *las arrugas* must sit in the applicative's complement position in (30a), which later becomes the grammatical subject via "Agree" (30b); these two constituents would enter a binding chain *à la* Cinque (1990), hence the identity relation they bear in terms of case and theta roles. Recomplementation data suggest that this is the case (Demonte and Fernández Soriano 2009; López 2009); Villa-García (2012, 2015) explains that clitic left-dislocated phrases that are sandwiched between two complementizers must be base generated, and that these phrases fail to show reconstruction effects, unlike their counterparts without recomplementation. In (31), the DP *las arrugas* appears sandwiched between two complementizers, which reinforces the idea that this constituent is left dislocated. Additionally, (31) shows that, should the negatively quantified dative DP be followed by a complementizer, the quantificational reading fails to obtain, proving that this position is indeed extrasentential; unlike the DP *las arrugas*, the quantified dative DP *a nadie* must sit in an A-position inside the sentence in (31), i.e., in Spec,TP.

(31) Dice que las arrugas, que a nadie (*que) se le ven fácilmente.
says that the wrinkles that nobody.DAT that RFL 3SG.DAT see easily
'He says that, the wrinkles, that nobody's, that they are easy to see.'

Notice, however, that when neither the dislocated theme DP nor the dislocated dative DP contain a quantifier, both can be sandwiched between complementizers.

(32) a. Dice que a Ismael, que las arrugas, (que) se le ven fácilmente.
says that Ismael.DAT that the wrinkles that RFL 3SG.DAT see easily
'He says that Ismael's wrinkles, that they are easy to see.'
b. Dice que las arrugas que a Ismael, (que) se le ven fácilmente.
says that the wrinkles that Ismael.DAT that RFL 3SG.DAT see easily
'He says that the wrinkles, that Ismael's, that they are easy to see.'

Additionally, it is well-known that CLLDs are sensitive to strong islands (e.g., complex NPs (33a) and adjuncts (33b)), but insensitive to weak islands (e.g., wh-islands (33c)) (Zubizarreta 1999; Bosque and Gutiérrez-Rexach 2009; Olarrea 2012, inter alia).

(33) a. *Complex NP island*
*Estoy convencido de que **a Paula$_i$** la enfermera conoce
I-am convinced of that Paula.ACC the nurse knows
a la doctora que **la$_i$** examinó.
to the doctor.ACC that her examined
'I am convinced that the nurse knows the doctor who examined Paula.'

b. *Adjunct island*
*Nos parece mejor que **a Paula$_i$** cocinemos la cena
1PL.DAT seems better that Paula.ACC we-cook the dinner
antes de avisar**la$_i$**.
before of tell-her
'We believe it is best we cook dinner before telling Paula.'

c. *Wh-island*
A Paula$_i$ no sé cómo podríamos averiguar quién
Paula.ACC not I-know how we-could guess who
la$_i$ invitó.
her invited
'Paula, I don't know how we could figure out who invited her.'

If preverbal subjects and dative DPs in middle-passive constructions are indeed instances of CLLDs, we should expect the same behavior regarding strong and weak islands; the data in (34) prove this hypothesis correct.

(34) *Complex NP island*
a. *Estoy seguro de que **las arrugas$_i$**, la doctora conoce
I-am sure of that the wrinkles the doctor knows
a la persona a la que se le ven ***pro*$_i$** fácilmente.
the person.ACC to the that RFL 3SG.DAT see pro easily
'I am sure that the doctor knows the person whose wrinkles are easily visible.'

b. *Estoy seguro de que, **a Ismael$_i$**, la doctora examinó
I-am sure of that Ismael.DAT the doctor examined
las arrugas que se **le$_i$** ven fácilmente.
the wrinkles.ACC that RFL 3SG.DAT see easily
'I am sure that the doctor examined Ismael's wrinkles that are easily visible.'

Adjunct island

c. ?Nos parece mejor que, **las arrugas$_i$**, cocinemos la cena
1PL.DAT seems better that the wrinkles.ACC we-cook the dinner
antes de que se le vean ***pro*$_i$** a Juan fácilmente.[14]
Before of that RFL 3SG.DAT see pro Juan.DAT easily
'We believe it is best we cook dinner before Juan's wrinkles are easily visible.'

d. *Nos parece mejor que, **a Juan$_i$**, cocinemos la cena
1PL.DAT seems better that Juan.DAT we-cook the dinner
antes de que se **le$_i$** vean las arrugas fácilmente.
before of that RFL 3SG.DAT see the wrinkles easily
'We believe it is best we cook dinner before Juan's wrinkles are easily visible.'

Wh-island

e. **A Ismael$_i$** no sé cómo podríamos averiguar si
Ismael.DAT not I-know how we-could guess if
se **le$_i$** ven las arrugas fácilmente.
RFL 3SG.DAT see the wrinkles easily
'As for Ismael, I don't know how we could figure out whether his wrinkles are easy to see.'

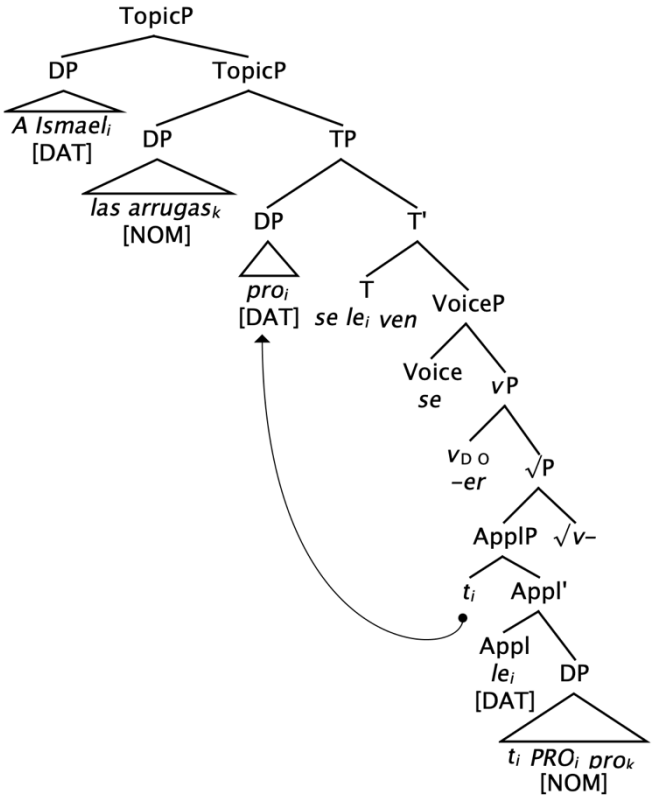
f. **Las arrugas$_{\mathbf{i}}$**, no sé cómo podríamos averiguar si
the wrinkles not I-know how we-could guess if
a Juan se le ven ***pro*$_{\mathbf{i}}$** fácilmente.
Juan.DAT RFL 3SG.DAT see pro easily
'As for Ismael's wrinkles, I don't know how we could figure out whether they are easy to see.'

g. **A Ismael$_{\mathbf{i}}$, las arrugas$_{\mathbf{k}}$**, no sé cómo podríamos averiguar
Ismael.DAT the wrinkles not I-know how we-could guess
si se **le$_{\mathbf{i}}$** ven ***pro*$_{k}$** fácilmente.
if RFL 3SG.DAT see easily
'As for Ismael, I don't know how we could figure out whether his wrinkles are easy to see.'

The middle-passive examples in (34), which contain dative DPs, mirror the contrasts of those in (33), involving CLLD configurations. Therefore, we can establish that middle-passive constructions with preverbal lexical DPs—dative or otherwise—are instances of CLLDs. I propose that the derivation of a middle-passive sentence where both a dative possessor DP and the theme DP appear preverbally is the following:

(35) a. A Ismael, las arrugas, se le ven fácilmente.

b.



Assuming Rizzi's (1997) cartographic approach to the sentence's left periphery, whereby left-dislocated constituents merge in recurring Topic projections outside the sentence, the dative possessor DP *a Ismael* and theme DP *las arrugas* in (35) are base generated extrasententially inside two different Topic projections. These two constituents corefer with two empty pronouns inside the sentence; the *pro* standing for the external possessor originates inside the possessum DP containing the *pro* coreferring with *las arrugas*, specifically in Spec,nP, from where it binds the PRO in Spec,NP. The possessor argument, unable to be case-marked inside the possessum DP, exits it and raises to the specifier of the low applicative head, where it licenses dative case. When T° looks down to probe a DP to check its EPP feature, it finds the possessor *pro* in Spec,ApplP first, for it is the closest one to T°, and it probes it to its specifier. Because the possessor argument is already case-marked, T° later assigns

nominative case to the null possessee via Agree, which also triggers verbal agreement. In this derivation, no intervention effects arise, minimality is respected, and the desired word order is obtained. Furthermore, this proposal harkens back to classic analyses contending that preverbal subjects in Spanish are CLLDs, such as Contreras (1976), Olarrea (1996), and Ordóñez and Treviño (1999), or Barbosa (1996) for European Portuguese.

Something is still to be said about middle-passive sentences with quantified DPs, like (36).

(36) Dice que a Ismael$_i$, que [ninguna arruga]$_k$ (*que) se le$_i$ ve fácilmente.
says that Ismael.DAT that none wrinkle that RFL 3SG.DAT see easily
'He says that none of Ismael's wrinkles is easily visible.'

In (36), the negatively quantified theme DP *ninguna arruga* appears to have raised to the preverbal subject position over the null pronoun standing for the external possessor in argument position, i.e., in Spec,ApplP; note that the theme's quantificational reading is obtained, and it cannot be followed by a complementizer. In other words, minimality seems to have been violated. Barbosa (2009) explains that there is a subset of quantificational expressions that are fronted by A'-movement without requiring contrastive focus, and this seems to be one of those cases. Therefore, I propose that the negatively quantified theme in (36) is fronted and adjoined to an A'-position above the null external possessor in Spec,TP, as shown in (37).

(37) Dice que **a Ismael$_i$**, que [$_{TP}$ [**ninguna arruga**]$_k$ [$_{TP}$ ***pro*$_i$** [$_T$ se le$_i$ ve t_k fácilmente]]]

Evidence for the fronting of this constituent via A'-movement comes from the fact that in Asturian, as well as in other Romance languages like European Portuguese (cf. Barbosa 2009), when these quantificational expressions appear preverbally, they trigger proclisis (38c), as in other contexts where A'-movement takes place; this is shown in (38d).

(38) a. Les engurries vénse-y fácil.
the wrinkles see.RFL-3SG.DAT easy
'His wrinkles are easily visible.'

b. A Ismael$_i$ vénse-y$_i$ les engurries fácil.
Ismael.DAT see.RFL-3SG.DAT the wrinkles easy
'Ismael's wrinkles are easily visible.'

c. Diz que a Ismael$_i$, que **nenguna engurria se-y$_i$** ve fácil.
says that Ismael.DAT that none wrinkle RFL-3SG.DAT see easy
'He says that none of Ismael's wrinkles is easily visible.'

d. A ISMAEL$_i$ se-y$_i$ ven les engurries fácil.
Ismael.DAT RFL-3SG.DAT see the wrinkles easy
'It is Ismael's wrinkles that are easily visible.'

In this section, I have shown evidence supporting the idea that preverbal lexical dative and theme DPs in Spanish middle-passive constructions are left-dislocated constituents coreferring with empty pronominals in argument position, i.e., CLLDs. This analysis avoids the intervention effects that a low applicative analysis of dative possessors in these structures would run into, if we assume that lexical DPs are generated inside the sentence, specifically, when accounting for the derivation where both the dative possessor DP and theme DP occur preverbally. In the next section, I will discuss how these data can be accounted for within a biclausal analysis of left dislocations (Ott 2014, 2015).

4.2. Biclausal/Paratactic Approach

Base generation and movement analyses of CLLDs face what some authors call Cinque's paradox (Cinque 1983, 1990; Ott 2014, 2015), i.e., while dislocated XPs are extrasentential constituents, in some respects, they behave as though they have moved from within the host clause. On the one hand, base-generation accounts must posit that the dislocated constituent and the resumptive element in CLLDs enter a special type of binding chain

in order to derive the connectivity between the two, as well as their sensitivity to strong islands. On the other, movement approaches must find answers for CLLDs' insensitivity to weak islands, lack of weak crossover effects, ability to license parasitic gaps, and the lack of subject–verb inversion in languages like Spanish (Ott 2014, 2015).

Recently, Ott (2014, 2015) elaborated a noncartographic analysis of left dislocations that appears to be able to do away with said paradox. This author claims that left-dislocated XPs are elliptical sentence fragments surfacing in linear juxtaposition to their host clause; in other words, dislocated constituents do not move to or are base generated in a left-peripheral projection. According to this author, two clauses are parallel, differing only in that CP1 contains Σ, i.e., the segment fragment, whereas the host clause contains K in its place, a free weak proform that is cross-sententially connected to Σ, thus enabling redundant material to delete. The biclausal representation of a Spanish CLLD is sketched in (39):

(39) [$_{CP1}$ ~~Ya leímos~~ [$_{\Sigma}$ **ese libro**]$_i$]
[$_{CP2}$ ya [$_{K}$ **lo**]$_i$ leímos]
'That book, we already read it.'

Technically, the two sentences in (39) are not syntactically connected, but doubly endophorically linked through cataphoric ellipsis and anaphoric K; semantically speaking, the second sentence is a reformulation of the first one. Finally, the fact that the dislocate constituent merges in the specifier of a left-peripheral projection in monoclausal analyses does not explain how this constituent is case-marked or how it receives its theta role; on the other hand, under a biclausal approach, Σ and K bear an identity relation because they enter identical case and theta relations in their respective clauses.

Because preverbal dative DPs and subject DPs in middle-passive sentences are left dislocated, a biclausal analysis of these configurations would involve three juxtaposed CPs, the third of which, the host CP, has two proforms, i.e., two Ks, one standing for the dative possessor and another one instantiating the theme argument, as shown in (40). The three CPs would account for the fact that each dislocated element may surface preceded by the complementizer *que* and in different intonational contours.

(40) [$_{CP1}$ (que) **[$_{\Sigma 1}$ a Ismael$_{DAT}$]$_i$** ~~se le ven las arrugas fácilmente~~]
[$_{CP2}$ (que) ~~a Ismael$_{DAT}$ se le ven~~ **[$_{\Sigma 2}$ las arrugas$_{NOM}$]$_k$** ~~fácilmente~~]
[$_{CP3}$ (que) **[$_{K1}$ *pro*$_{DAT}$]$_i$** se le ven **[$_{K2}$ *pro*$_{NOM}$]$_k$** fácilmente]

Alternatively, cases where the theme and dative DPs surface preverbally, the former preceding the latter, would be derived as follows:

(41) [$_{CP1}$ (que) ~~a Ismael$_{DAT}$ se le ven~~ **[$_{\Sigma 2}$ las arrugas$_{NOM}$]$_k$** ~~fácilmente~~]
[$_{CP2}$ (que) **[$_{\Sigma 1}$ a Ismael$_{DAT}$]$_i$** ~~se le ven las arrugas fácilmente~~]
[$_{CP3}$ (que) **[$_{K1}$ *pro*$_{DAT}$]$_i$** se le ven **[$_{K2}$ *pro*$_{NOM}$]$_k$** fácilmente]

The derivations in (40) and (41) still account for the fact that the dative possessor—be it a full DP or an empty pronominal—is the structurally higher constituent, and therefore, the closest one to T°, for it sits in the specifier of ApplP, whereas the theme DP merges in Appl's complement position. Consequently, T° will always probe the dative possessor to its specifier, never the theme DP, thus, avoiding any potential minimality violation.

5. Conclusions

In this paper, I have examined data from Spanish middle-passive constructions containing dative possessor DPs. These arguments originate inside the possessum DP and raise to the specifier of a low applicative of possession head to be case-licensed. I showed that an applicative analysis of these structures runs into a minimality violation when accounting for the configuration where both the dative possessor DP and the theme DP occur preverbally, if we assume that the former is left dislocated and the latter sits in the preverbal subject position: here, Spec,TP. When T° looks down to probe a DP to its specifier, it would have to

skip an empty pronominal in SpecApplP standing for the possessor, in order to attract the theme DP in argument position.

(42) **a Ismael**$_i$ [$_{TP}$ las arrugas$_k$ [$_T$ se le$_i$ ven [$_{VoiceP}$ [$_{vP}$ [$_{√P}$ [$_{ApplP}$ ***pro***$_i$ [$_{Appl}$ ~~le~~ [$_{DP}$ ***t***$_k$]]] ~~√v~~]]]]]

Instead, I have provided evidence supporting the idea that preverbal lexical dative and theme DPs are CLLDs coreferring with empty pronominals in argument position. Thus, the null dative possessor pronoun in Spec,ApplP will always raise to Spec,TP, while the empty theme pronoun will remain inside the verbal domain, therefore avoiding any intervention effects.

(43) **[a Ismael$_i$]**[**las arrugas$_k$**][$_{TP}$ *pro*$_i$ [$_T$ se le$_i$ ven [$_{VoiceP}$ [$_{vP}$ [$_{√P}$ [$_{ApplP}$ ***t***$_i$ [$_{Appl}$ ~~le~~ [$_{DP}$ ***pro***$_k$]]] ~~√v~~]]]]]

Moreover, the fact that these constituents are extrasentential straightforwardly derives their unmarked preverbal position and their aforementioned status as sentential topics. Finally, I have also explored how these data can be successfully analyzed within a biclausal analysis of left dislocations, whereby left-dislocated XPs are elliptical sentence fragments surfacing in linear juxtaposition to their host clause. These constituents do not move to or are base generated in a left-peripheral projection; instead, two clauses are parallel, differing only in that CP1 contains Σ, i.e., the segment fragment, whereas the host clause contains K in its place, a free weak proform that is cross-sententially connected to Σ, thus enabling redundant material to delete. This type of approach also avoids any minimality violation.

(44) [$_{CP1}$ (que) **[Σ1 a Ismael$_{DAT}$]$_i$** ~~se le ven las arrugas fácilmente~~]
[$_{CP2}$ (que) ~~a Ismael$_{DAT}$ se le ven~~ **[Σ2 las arrugas$_{NOM}$]$_k$** ~~fácilmente~~]
[$_{CP3}$ (que) **[K1 *pro*$_{DAT}$]$_i$** se le ven **[K2 *pro*$_{NOM}$]$_k$** fácilmente]

The data presented in this study contribute to the much-debated position that preverbal subjects and dative DPs occupy in Spanish, aligning with classic proposals by Contreras (1976), Olarrea (1996), and Ordóñez and Treviño (1999), who claim that Spanish preverbal subjects are left-dislocated constituents. Similarly, Barbosa (1996) arrived at the same conclusion for preverbal subjects in European Portuguese. Moreover, these results also support Baker's (1995) intuitions for polysynthetic languages, where lexical DPs merge extrasententially and are coindexed with empty pronominals in argument position. Finally, further research is still required to establish whether this is a common trend in other Romance languages as well; I leave this question open for future inquiry.

Funding: This research was funded by an award from the Humanities Scholarship Enhancement Fund from the College of Liberal Arts and Sciences at the University of Florida.

Acknowledgments: I am thankful to three anonymous reviewers whose expertise contributed to improve this manuscript. Additionally, I would like to thank Ana Ojea, Antxon Olarrea, Eric Potsdam and Julio Villa-García for their incredibly useful comments on these data and analysis. Finally, I would also like to express my gratitude to the conference abstract reviewers and audience at the 53rd Linguistic Symposium on Romance Languages in Paris, as well as to the organizers and audience of the Romance Linguistics Circle, at the Universities of Cambridge and Newcastle, for their valuable feedback on this project. All errors are exclusively my own.

Conflicts of Interest: The author declares no conflict of interest.

Notes

1 Excluding a habitual *se*-passive interpretation, i.e., *these telephones are usually repaired easily.*

2 Cardinaletti (1998) classifies Romance possessive pronouns as clitic (i), weak (ii), or strong (iii):

(i) Mi teléfono. *Spanish*
(ii) El meu telèfon. *Catalan*
(iii) El teléfono mío *Spanish*
'My telephone.'

3 Although dative possessors are highly frequent with body-part nouns, external possession can also involve alienably possessed entities (cf. Cuervo 2003, ex. 87). In addition to the possessor interpretation, the dative may receive an additional connotation, depending on the verb's semantics, such as affected, benefactive, source, goal, etc.

(i) Le_i vi/limpié $[\text{la habitación}]_i$ a Miguel_i.
3SG.DAT saw/cleaned the room Miguel.DAT
'I saw/cleaned Miguel's room.'

4 It has been noted, however, that dative possessors are compatible with possessive determiners in some Latin American varieties of Spanish in contact with indigenous languages; see Escobar (1992), Rodríguez-Mondoñedo (2019), and Giancaspro and Sánchez (2021), among others, for details.

5 This idea harkens back to Guéron (1985), who initially proposed that inalienably possessed nouns take an empty category of some sort as an argument in their specifier.

6 Example from Catalan.

7 Because internal possession is not the focus of this paper, I will not describe Suárez-Palma's proposal in depth; I refer the reader to that work for details on how this phenomenon comes about.

8 In other Romance languages like Catalan, possessives and dative possessors are also in a complementary distribution:

(i) Les (*meves) arrugues s'em veuent fàcilment.
the my wrinkles RFL-1SG.DAT see easily
'My wrinkles are easily visible.'

9 In order to account for the fact that preverbal subjects cannot be bare NPs in Spanish, Suñer (1982, p. 209) proposed the Naked Noun Phrase Constraint.

(i) The Naked Noun Phrase Constraint: 'An unmodified common noun in preverbal position cannot be the surface subject of a sentence under conditions of normal stress and intonation.'

10 Excluding a habitual passive reading, i.e., *these blouses are usually washed easily.*

11 Sánchez López (2002) points out that modification is not always required in Spanish when the event refers to a property that defines the notional object as a particular type, or in the context of negation, for it can trigger genericity by negating the possibility operator, therefore denoting the absence of a given property.

(i) Este fruto (no) se come.
this fruit not RFL eats
'This fruit is (not) edible.'

12 The fact that quantifiers cannot be dislocated was also observed in Cinque (1990), Dobrovie-Sorin (1990), Rizzi (1997), Barbosa (2000), and Arregi (2003).

13 For movement analyses of left dislocation, see Cinque (1977), Dobrovie-Sorin (1990), Kayne (1994), Villalba (2000), and López (2003), to name a few.

14 The different degree of acceptability between (34c) and (34d) has to do with the fact that the former is ambiguous and can also be interpreted as a hanging topic configuration. Hanging topics do not bear an identity relationship with a resumptive element in the host clause (cf. Villa-García and Ott 2022); the lack of subject clitic pronouns in Spanish favors this ambiguity. However, if both the dislocated theme and resumptive element were marked with accusative case, the ungrammaticality becomes clearer.

(i) Nos parece mejor que, las arrugas_i, cocinemos la cena
1PL.DAT seem better that the wrinkles.ACC we-cook the dinner
antes de vérse**las$_i$** a Juan.
before of see.3SG.DAT.3PL.ACC Juan.DAT
'We believe it is best we cook dinner before we see Juan's wrinkles.'

References

Ackema, Peter, and Maaike Schoorlemmer. 1995. Middles and nonmovement. *Linguistic Inquiry* 26: 172–97.

Ackema, Peter, and Maaike Schoorlemmer. 2006. Middles. In *The Blackwell Companion to Syntax*. Edited by Martin Everaert and Henk van Riemsdijk. Oxford: Blackwell Publishing Ltd., vol. I, pp. 131–203.

Alexiadou, Artemis, Liliane Haegeman, and Melita Stavrou. 2007. *Noun Phrase in the Generative Perspective*. New York: Mouton de Gruyter.

Armstrong, Grant. 2021. Te puse la mano en el hombro, 'I put my hand on your shoulder.' A solution to a puzzling constraint on multiple external possession relations in Spanish. *Isogloss: Open Journal of Romance Linguistics* 7: 1–28.
Arregi, Karlos. 2003. Clitic left dislocation is contrastive topicalization. *University of Pennsylvania Working Papers in Linguistics* 9: 31–44.
Baker, Mark. 1995. *The Polysynthesis Parameter*. Oxford: Oxford University Press.
Bally, Charles. 1926. L'expression des idées de sphere personnelle et de solidarité dans les langues indo-européennes. In *Festschrift Louis Gauchat*. Edited by Frantz Frankhauser and Jakob Jud. Arau: Sauerländer, pp. 68–78.
Barbosa, Pilar. 1996. Clitic placement in European Portuguese and the position of subjects. In *Approaching Second: Second Position Clitics and Related Phenomena*. Edited by Aaron Halpern and Arnold Zwicky. Stanford: CSLI, pp. 1–41.
Barbosa, Pilar. 2000. Clitics: A window into the null-subject property. In *Essays in Portuguese Comparative Syntax*. Edited by João Costa. Oxford: Oxford University Press, pp. 31–93.
Barbosa, Pilar. 2009. Two kinds of subject *pro*. *Studia Linguistica* 63: 2–38. [CrossRef]
Bosque, Ignacio, and Javier Gutiérrez-Rexach. 2009. *Fundamentos de Sintaxis Formal*. Madrid: Akal.
Cardinaletti, Anna. 1998. On the deficient/strong opposition in possessive systems. In *Possessors, Predicates and Movement in the Determiner Phrase*. Edited by Artemis Alexiadou and Chris Wilder. Amsterdam: John Benjamins Publishing, pp. 17–53.
Chierchia, Gennaro. 2003. A semantics for unaccusatives and its syntactic consequences. In *The Unaccusativity Puzzle. Explorations of the Syntax-Lexicon Interface*. Edited by Artemis Alexiadou, Elena Anagnostopoulou and Martin Everaert. Oxford: Oxford University Press, pp. 60–84.
Chomsky, Noam. 1995. *The Minimalist Program*. Cambridge, MA: MIT Press.
Cinque, Guglielmo. 1977. The movement nature of left dislocation. *Linguistic Inquiry* 8: 397–412.
Cinque, Guglielmo. 1983. 'Topic' constructions in some European languages 'connectedness'. In *Connectedness in Sentence, Discourse and Text*. Edited by Konrad Ehlich and Henk van Riemsdijk. Tilburg: KUB, pp. 7–14.
Cinque, Guglielmo. 1990. *Types of A-Bar Dependencies*. Cambridge, MA: MIT Press.
Condoravdi, Cleo. 1989. The middle: Where semantics and morphology meet. *MIT Working Papers in Linguistics* 11: 18–30.
Conti, Carmen. 2011. Possessive dative revisited: Another view of external possession in Spanish. *Studia Linguistica* 65: 170–197. [CrossRef]
Contreras, Heles. 1976. *A Theory of Word Order with Special Reference to Spanish*. Amsterdam: North-Holland.
Cuervo, María Cristina. 2003. *Datives at Large*. Cambridge, MA: Massachusetts Institute of Technology dissertation.
Demonte, Violeta. 1988. El artículo en lugar del posesivo y el control de los sintagmas nominales. *Nueva Revista de Filología Hispánica* 36: 89–108. [CrossRef]
Demonte, Violeta. 1995. Dative alternation in Spanish. *Probus* 7: 5–30. [CrossRef]
Demonte, Violeta, and Olga Fernández Soriano. 2009. Force and finiteness in the Spanish complementizer system. *Probus* 21: 23–49. [CrossRef]
Dixon, Robert M. W. 1982. *Where Have All the Adjectives Fone? And Other Essays in Semantics and Syntax*. Berlin: Mouton.
Dobrovie-Sorin, Carmen. 1990. Clitic-doubling, *wh*-movement and quantification in Romanian. *Linguistic Inquiry* 21: 351–97.
Escobar, Anna María. 1992. El español andino y el español bilingüe. Semejanzas y diferencias en el uso del posesivo. *Lexis* 16: 189–222. [CrossRef]
Fábregas, Antonio. 2021. SE in Spanish: Properties, structures and analyses. *Borealis. An International Journal of Hispanic Linguistics* 10: 1–236. [CrossRef]
Fábregas, Antonio, Ángel Jiménez-Fernández, and Mercedes Tubino. 2017. What's up with dative experiencers? In *Romance Languages and Linguistic Theory 12*. Edited by Ruth E. V. Lopes, Juanito Ornelas de Avelar and Sonia Cyrino. Amsterdam: John Benjamins Publishing, pp. 29–48.
Fagan, Sarah. 1992. *The Syntax and Semantics of Middle Constructions. A Study with Special Reference to German*. Cambridge: Cambridge University Press.
Fernández Soriano, Olga. 1999. Two types of impersonal sentences in Spanish: Locative and dative subjects. *Syntax* 2: 101–40. [CrossRef]
Fernández Soriano, Olga, and Amaya Mendikoetxea. 2013. Non selected dative arguments in Spanish anticausative constructions. *The Diachronic Typology of Non-Canonical Subjects* 140: 3–34.
Fernández-Sánchez, Javier, and Dennis Ott. 2020. Dislocations. *Language and Linguistics Compass* 14: 1–39. [CrossRef]
Fodor, Janet Dean, and Ivan A. Sag. 1982. Referential and quantificational indefinites. *Linguistics and Philosophy* 5: 355–98. [CrossRef]
Frascarelli, Mara. 1997. The phonology of focus and topic in Italian. *The Linguistic Review* 14: 221–48. [CrossRef]
Frascarelli, Mara. 2000. *The Syntax-Phonology Interface in Focus and Topic Constructions in Italian*. Dordrecht: Kluwer.
Giancaspro, David, and Liliana Sánchez. 2021. *Me, mi, my*. Innovation and variability in heritage speakers' knowledge of inalienable possession. *Glossa: A Journal of General Linguistics* 6: 1–28. [CrossRef]
Guéron, Jacqueline. 1983. L'emploi possessif de l'article define en français. *Langue Française* 58: 23–55. [CrossRef]
Guéron, Jacqueline. 1985. Inalienable possession. PRO-inclusion and lexical chains. In *Grammatical Representation*. Edited by Jacqueline Guéron, Hans Georg Obenauer and Jean-Yves Pollock. Dordrecht: Foris, pp. 42–86.
Guéron, Jacqueline. 2006. Inalienable possession. In *The Blackwell Companion to Syntax*. Edited by Martin Everaert and Henk van Riemsdijk. Oxford: Blackwell Publishing Ltd., vol. I, pp. 589–638.

Gutiérrez Ordóñez, Salvador. 1999. Los dativos. In *Gramática Descriptiva de la Lengua Española*. Edited by Ignacio Bosque and Violeta Demonte. Madrid: Espasa Calpe, pp. 1855–1930.
Hale, Ken, and Samuel J. Keyser. 1988. Explaining and constraining the English middle. *Lexicon Project Working Papers* 24: 41–57.
Jiménez-Fernández, Ángel. 2020. The information structure of high and low datives and their psychological import. In *Beyond Emotions in Language. Psychological Verbs at the Interfaces*. Edited by Bozena Rozwadowska and Anna Bondaruk. Amsterdam: John Benjamins Publishing, pp. 213–44.
Kayne, Richard. 1994. *The Antisymmetry of Syntax*. Cambridge, MA: MIT Press.
Kempchinsky, Paula. 1992. The Spanish possessive dative construction: θ-role assignment and proper government. In *Romance Languages and Modern Linguistic Theory. Current Issues in Linguistic Theory*. Edited by Paul Hirschbúhler and E.F.K. Koerner. Amsterdam: John Benjamins Publishing, pp. 135–49.
Keyser, Samuel J., and Thomas Roeper. 1984. On the middle and ergative constructions in English. *Linguistic Inquiry* 3: 381–416.
Kliffer, Michael D. 1983. Beyond syntax: Spanish inalienable possession. *Linguistics* 21: 759–94. [CrossRef]
Kratzer, Angelika. 1996. Severing the external argument from its verb. In *Phrase Structure and the Lexicon*. Edited by Johan Rooryck and Laurie Zaring. Dordrecht: Springer, pp. 109–37.
Lekakou, Maria. 2005. In the Middle Somewhat Elevated. The Semantics of Middles and Its Crosslinguistic Realization. Ph.D. dissertation, University College London, London, UK.
López, Luis. 2003. Steps for a well-adjusted dislocation. *Studia Linguistica* 57: 193–231. [CrossRef]
López, Luis. 2009. *A Derivational Syntax for Information Structure*. Oxford: Oxford University Press.
Masullo, Pascual. 1992. Quirky datives in Spanish and the non-nominative subject parameter. *MIT Working Papers in Linguistics* 16: 82–103.
Mendikoetxea, Amaya. 1999. Construcciones con "se": Medias, pasivas e impersonales. In *Gramática Descriptiva de la Lengua Española*. Edited by Ignacio Bosque and Violeta Demonte. Madrid: Espasa Calpe, pp. 1575–1630.
Nakamoto, Takeshi. 2010. Inalienable possession constructions in French. *Lingua* 120: 74–102. [CrossRef]
O'Grady, William D. 1980. The derived intransitive construction in English. *Lingua* 52: 57–72. [CrossRef]
Olarrea, Antxon. 1996. Pre- and Postverbal Subjects in Spanish: A Minimalist Account. Ph.D. dissertation, University of Washington, Seattle, WA, USA.
Olarrea, Antxon. 2012. Word order and information structure. In *The Handbook of Hispanic Linguistics*. Edited by José Ignacio Hualde, Antxon Olarrea and Erin O'Rourke. Oxford: Wiley-Blackwell.
Ordóñez, Francisco, and Esthela Treviño. 1999. Left dislocated subjects and the pro-drop parameter: A case study of Spanish. *Lingua* 107: 39–68. [CrossRef]
Ott, Dennis. 2014. An ellipsis approach to contrastive left-dislocation. *Linguistic Inquiry* 45: 269–303. [CrossRef]
Ott, Dennis. 2015. Connectivity in left-dislocation and the composition of the left periphery. *Linguistic Variation* 15: 225–90. [CrossRef]
Picallo, M. Carme, and Gemma Rigau. 1999. El posesivo y las relaciones posesivas. In *Gramática Descriptiva de la Lengua Española*. Edited by Ignacio Bosque and Violeta Demonte. Madrid: Espasa Calpe, pp. 973–1024.
Pylkkänen, Liina. 2002. Introducing Arguments. Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge, MA, USA.
Rigau, Gemma. 1988. Strong Pronouns. *Linguistic Inquiry* 19: 503–11.
Rivero, María Luisa. 1980. On left-dislocation and topicalization in Spanish. *Linguistic Inquiry* 11: 363–93.
Rizzi, Luigi. 1990. *Relativized Minimality*. Cambridge, MA: The MIT Press.
Rizzi, Luigi. 1997. The fine structure of the left periphery. In *Elements of Grammar*. Edited by Liliane Haegeman. Dordrecht: Kluwer, pp. 281–337.
Rizzi, Luigi. 2012. Minimality. In *The Oxford Handbook of Linguistic Minimalism*. Edited by Cedric Boeckx. Oxford: Oxford University Press, pp. 220–38.
Roberts, Ian. 1987. *The Representation of Implicit and Dethematized Subjects*. Dordrecht: Foris.
Rodrigues, Cilene. 2010. Possessor raising through thematic positions. In *Movement Theory of Control*. Edited by Norbert Hornstein and Maria Polinsky. Amsterdam: John Benjamins Publishing, pp. 119–46.
Rodrigues, Cilene. 2023. Possessor raising and structural variations within the *v*P-domain. *Glossa: A Journal of General Linguistics* 8: 1–29. [CrossRef]
Rodríguez-Mondoñedo, Miguel. 2019. Syntactic phenomena in Peruvian Spanish. In *The Syntactic Variation of Spanish Dialects*. Edited by Ángel Gallego. Oxford: Oxford University Press, pp. 245–74.
Sánchez López, Cristina. 2002. Las construcciones con *se*. Estado de la cuestión. In *Las Construcciones Con se*. Edited by Ángel Cristina Sánchez López. Madrid: Visor Libros, pp. 18–167.
Sánchez López, Cristina. 2007. The possessive dative and the syntax of affected arguments. *Cuadernos de lingüística del Instituto Universitario de Investigación Ortega y Gasset* 14: 153–73.
Schäfer, Florian. 2008. *The Syntax of (Anti-)causatives: External Arguments in Change-of-State Contexts*. Amsterdam: John Benjamins Publishing, p. 126.
Stroik, Thomas. 1992. Middles and movement. *Linguistic Inquiry* 23: 127–37.
Suárez-Palma, Imanol. 2019. Stuck in the Middle: Dative Arguments and Middle-Passive Constructions in Spanish. Ph.D. dissertation, University of Arizona, Tucson, AZ, USA.

Suárez-Palma, Imanol. 2020. Applied arguments in Spanish inchoative middle constructions. *Glossa: A Journal of General Linguistics* 5: 1–37. [CrossRef]
Suárez-Palma, Imanol. 2022. Middle formation and inalienability in Asturian. In *Sound, Syntax and Contact in the Languages of Asturias*. Edited by Guillermo Lorenzo. Amsterdam: John Benjamins Publishing, pp. 131–49.
Suárez-Palma, Imanol. Forthcoming. Inalienable possession (and lack thereof) in Spanish. *Studies in Hispanic and Lusophone Linguistics*. *in press*.
Suñer, Margarita. 1982. *Syntax and Semantics of Spanish Presentational Sentence-Types*. Washington, DC: Georgetown University Press.
Tubino, Mercedes. 2007. Preverbal Datives in Spanish. How Quirky Are They? An Agree Analysis of Gustar. Master's dissertation, University of Arizona, Tucson, AZ, USA.
Vergnaud, Jean Roger, and María Luisa Zubizarreta. 1992. The definite determiner and the. inalienable constructions in French and in English. *Linguistic Inquiry* 23: 595–652.
Villa-García, Julio. 2012. Recomplementation and locality of movement in Spanish. *Probus* 24: 257–314. [CrossRef]
Villa-García, Julio. 2015. *The Syntax of Multiple-que Sentences in Spanish: Along the Left Periphery*. Issues in Hispanic and Lusophone Linguistics. Amsterdam: John Benjamins Publishing.
Villa-García, Julio, and Dennis Ott. 2022. Recomplementation as a paratactic phenomenon: Evidence from Spanish and English. *Journal of Linguistics*, 1–32. [CrossRef]
Villalba, Xavier. 2000. The Syntax of Sentence Periphery. Ph.D. dissertation, Universitat Autònoma de Barcelona, Barcelona, Spain.
Zubizarreta, María Luisa. 1999. Las funciones informativas: Tema y foco. In *Gramática Descriptiva de la Lengua Española*. Edited by Ignacio Bosque and Violeta Demonte. Madrid: Espasa, pp. 4215–4244.

languages

MDPI

Article

C° realizations along the left edge across English and Spanish

Julio Villa-García [1,2]

[1] Department of English, French and German Philology, University of Oviedo, 33011 Oviedo, Principality of Asturias, Spain; julio.villa-garcia@manchester.ac.uk
[2] Department of Linguistics and English Language, The University of Manchester, Manchester M13 9PL, UK

Abstract: This paper investigates the lexicalization of the complementizer *that/que* in English and Spanish varieties in different contexts along the left edge of the clause. This is performed through discussion of a range of constructions traditionally attributed to the CP domain/left periphery, primarily (but not only) in certain embedded clauses. The ubiquity of *that/que*, that is, the lexical realization of *that/que* in subordinating environments, exclamative clauses, interrogative contexts, and subjunctive clauses, amongst others, sheds light not only on the characterization of the relevant constructions but also on the make-up of the left edge of the clause. The fact that such realizations can be obligatory, optional, or, on occasion, impossible, sometimes depending on the variety in question, furthers our understanding of head lexicalizations while contributing to macro and microvariation studies in syntactic theory. In so doing, this paper paves the way for holistic investigations devoted to complementizer realization in the head position of different left-edge-related constructions and in different linguistic varieties.

Keywords: left periphery; head lexicalizations; subordination; interrogatives; exclamatives; dislocations; finiteness; dialectal variation

Citation: Villa-García, Julio. 2023. C° realizations along the left edge across English and Spanish. *Languages* 8: 268. https://doi.org/10.3390/languages8040268

Academic Editors: Anna Cardinaletti and Ana Ojea

Received: 21 June 2023
Revised: 28 September 2023
Accepted: 27 October 2023
Published: 14 November 2023

1. Introduction

Complementizers offer a valuable window into the architecture of the leftmost part of clauses, the demarcation of the limits between the left edge (Complementizer Phrase, CP) and the inflectional/tense domain (IP/TP), and the analysis of a range of constructions traditionally attributed to the leftmost portion of the clause (i.e., the left periphery).

Beyond merely heralding an upcoming subordinate clause, as in *I think* [*that complementizers are mysteriously fascinating*], recent research has unearthed naturalistic data pointing to the conclusion that complementizers in languages like English and Spanish can occupy a host of positions along the clausal left-peripheral spine, contingent on the specific constructions at issue, as well as on the root vs. embedded dichotomy, in different varieties of said languages. Radford (2018), for instance, shows through numerous examples from spontaneous speech that spoken English is replete with what seems like different instances of *that*, boldfaced examples of which appear in (1):

(1) a. *I think* ***that*** *you are nice.*
b. *They told me* ***that*** *given the current crisis,* ***that*** *the building will remain closed.*
c. *Obviously* ***that*** *the Achilles was giving him a bit of a problem* (Ian Chappell, BBC Radio 5, cited in Radford (2018, p. 162)).
d. *I am not sure what kind of ban* ***that*** *FIFA has in mind* (Bert Millichip, BBC Radio 4, cited in Radford (2018, p. 159)).
e. *What a mine of useless information* ***that*** *I am!* (Sir Terry Wogan, BBC Radio 2, cited in Radford (2018, p. 159)).
f *Please, ensure* ***that*** *if your faculty commit to permitting candidates to attend their classes,* ***that*** *there be sufficient diversity of courses and* ***that*** *syllabi permit visitors to attend* (Official university communication, Pennsylvania, 20 November 2013, cited in Villa-García (2015, p. 96)).

In (1)a, we observe a classical example of subordinating *that*, which can typically remain silent in non-formal contexts. (1)b illustrates the double-*that* (cf. recomplementation) configuration involving a sandwiched element flanked by overt instances of the complementizer (Villa-García 2015, 2019; Radford 2018; Villa-García and Ott 2022, among others). (1)c displays an initial adverb followed by *that*. (1)d and (1)e, respectively, feature interrogative and exclamative phrases followed by an instance of *that*. Lastly, (1)f exemplifies an instance of a secondary *that* termed 'jussive/optative' *that* by Villa-García (2015). This complementizer, which shares much in common with the recomplementation *that* (cf. (1)b), as we shall see, is associated with the subjunctive mood. Note that in all the cases in (1), the different occurrences of *that* can in principle be silent for most speakers without obvious semantic consequences, an issue to which we return in due course.

Sentences featuring putatively different *que*s are indeed also attested in (certain varieties of) Spanish (see, for a subset of cases, Villa-García (2015)), as indicated by the data in (2). Note that not all sentences are attested in all varieties of the language.

(2) a. *Cree* ***que*** *llueve*
believes that rains.
'S/He believes (that) it's raining.'

b. *Me* *contaron* ***que*** *a* *María,* ***que*** *no* *la* *llaman.*
cl. told that acc. Mary that not cl. call
'They told me that Mary, that they are not calling her.'

c. *Ahí* *sí* ***que*** *no* *voy.*
there yes that not go
'I'm certainly not going there.'

d. ¿ *Por* *qué* ***que*** *viniste?* (Diego Gibanal Faro, pers. comm. 2023)
for what that came
'Why did you come (here)?'

e. ¡ *Qué* *alto* ***que*** *eres!*
what tall that are
'How tall you are!'

f. ¡ *Vaya* ***que*** *si* *voy!*
vaya that if go
'Of course I am going!/How can you even wonder if I'm going?'

g. ¡ *A* *Madrid* ***que*** *me* *piro!*
to Madrid that cl. *piro!*
'I am off to Madrid!'

h. Por supuesto ***que*** no me quedo.
of course that not cl. stay
'Of course I am not staying!'

i. ¡ *Ojalá* ***que*** *ganemos* *Eurovisión!*
God-willing that $\text{win}_{\text{Subjunctive}}$ Eurovision
'I hope we win the Eurovision contest.'

j. *Gritó* ***que*** *a* *la* *fiesta,* ***que*** *vaya* *Marta.*
shouted that to the party that $\text{go}_{\text{Subjuntive}}$ Marta
'S/He demanded by shouting that Martha go to the party.'

The Spanish paradigm displays cases akin to the English ones, including the subordinating complementizer, (1)a and (2)a, recomplementation cases, (1)b and (2)b, interrogative and exclamative complementizers, (1)d,e and (2)d,e, adverbial cases, (1)c and (2)h, and 'jussive/optative' cases, (1)f and (2)j. Additionally, the Spanish paradigm also includes *sí (que)* 'yes that' cases, (2)c, investigated from the standpoint of microvariation by Villa-García and Rodríguez (2020); (2)f, which shows the exclamative particle *vaya* plus *que* plus *si* 'if/whether'; (2)g, the emphatic construction involving a fronted constituent plus *que* plus sentence; and (2)i, which instantiates the desiderative/optative construction with *ojalá* (from the Hispanic Arabic expression *law šá lláh* 'if God wants') plus *que* plus a verb in the subjunctive.

Multiple-complementizer sentences like the ones in (1) and (2) raise several intriguing questions, such as:

(i) Is *that/que* a mere overt manifestation of an otherwise null/silent head (i.e., [$_{XP}$ YP [$_{X'}$ *that/que* vs. Ø . . .]])? If so, is *that/que* always present but deleted in the phonology (PF), ~~*that/que*~~, or else inserted when phonologically realized?
(ii) Is the presence of *that/que* indicative of a more complex underlying structure instead?
(iii) Whatever the case may be, what determines the (non-)realization of the complementizer in different positions? Is it dialect-based? If so, are some dialects more prone to lexicalizing complementizers in different positions than others? Is complementizer lexicalization processing-based? Are there any other factors that play a role in determining the presence or absence of the complementizer, such as formal vs. informal contexts?
(iv) Does the presence vs. absence of *that/que* have a bearing on the semantics (LF)? Put another way, is complementizer realization just a PF matter, or are there LF reflexes as well?

Set against this background, the present paper aims to constitute a first step towards partially answering the questions in (i)–(iv), in the hope that future studies will further advance our understanding of the puzzle presented herein. In the course of the ensuing discussion, additional questions will be posed which are relevant to various ongoing debates in syntax. Similarly, old and new data will be provided throughout, and previously unnoticed dialectal contrasts will be brought to light. In this sense, therefore, this paper aims to contribute to the diatopic mapping of the relevant constructions. This paper is thus meant to make both a descriptive and a theoretical contribution.

This paper is organized as follows: Section 2 presents existing accounts of the clausal left edge within the transformational generative paradigm; Section 3, which constitutes the bulk of the paper, returns to the data above and discusses a subset of the constructions in turn, in both English and Spanish, though mention of other linguistic varieties will be made when appropriate. In this connection, I concentrate on four major types of complementizers shared by English and Spanish, namely high complementizers, topic-related complementizers (cf. recomplementation), focal complementizers (exclamatives and questions), and low complementizers ('jussive/optative' complementizers). A critical review of existing accounts of the particular constructions is provided for each case; Section 4 turns to general extant accounts of inter- and intra-linguistic variation in relation to the presence vs. absence of complementizers and makes new analytical suggestions; and Section 5 is the conclusion.

2. The Analysis of the Leftmost Part of Sentences in the Generative Tradition

A complementizer such as *that/que* is assumed in the Chomskyan tradition to be a (functional) head, as shown abstractly in (3). I will take this conception of complementizers as heads as my point of departure (though see Vincent (2019) and references therein for a skeptical view).

(3) [$_{XP}$ [$_{X'}$ *that/que*]]

Since the seminal work of *Barriers* (Chomsky 1986), complementizers have by and large been taken to occupy the head of CP, as follows:

(4) [$_{CP}$ [$_{C'}$ *that/que*]]

Data along the lines of some of the sentences in (1) and (2) clearly indicate that a single CP projection may not suffice, as multiple left-edge-related constructions may occur concurrently in a given sentence/clause simultaneously. This in fact led to the postulation of CP recursion, namely the ability of multiple CPs to occur in a given clause (cf. (5)).

(5) [$_{CP}$ [$_{C'}$ [$_{CP}$ [$_{C'}$ [$_{CP}$ [$_{C'}$]]]]]]

Sentences displaying multiple instances of *that/que*, as in (1)b,f and (2)b,j appear to be ideal candidates for an analysis like that in (5), which is, in fact, the account pursued in Iatridou and Kroch (1992) for the English recomplementation case:

(6) [$_{CP}$ [$_{C'}$ *that/que* [$_{CP}$ XP [$_{C'}$ *that/que*]]]]

The late 1990s witnessed the emergence of cartographic approaches to the leftmost portion of clauses: Rizzi (1997) made the highly influential claim that there is a fine structure of the left-periphery (referred to as a templatic structure by those arguing against it ever since), which in effect splits the old CP domain into dedicated projections, each of which is devoted to hosting different elements, which Rizzi argues display the relative order in (7) (note that in Rizzi's system, TopicP is recursive in that multiple TopicP projections are allowed in the left-peripheral spine, if required by the presence of multiple topical phrases):[1]

(7) [ForceP [Force′ [TopicP [Topic′ [FocusP [Focus′ [FinitenessP [Finiteness′ IP/TP]]]]]]]]

CP

This type of account is rather appealing at first sight, and in fact has been assumed in much work into a broad array of languages since then. For instance, on this view, high complementizers spell out Force°; dislocated phrases (such as Clitic Left Dislocations, ClLDs) occupy the specifier of TopicP; wh-phrases and exclamative phrases (which are mutually exclusive) are the occupants of Spec, FocusP; low complementizers and elements related to mood and finiteness are associated with FinitenessP. For Rizzi (1997), TopicP and FocusP are only projected on an as-needed basis (that is, when topical or focal elements occur), and only the former is recursive, as noted. As Rizzi (2013, p. 200) observes, "the left periphery is populated by a system of functional heads dedicated to the expression of scope-discourse properties", thus:

(8) a. *Which car* Q *should I purchase?*
b. *This promise of lifelong service* TOP *I renew to all today* (King Charles III, cited in Villa-García (2023, p. 2)).
c. *THIS BOOK* FOC *you should buy (, not that one).*
d. *What an incredible week* EXCL *we have had!*

Accordingly, an exclamatory sentence featuring an exclamative phrase such as (8)d would receive the following simplified analysis under Rizzi's approach (note that it is immaterial to the present discussion whether exclamatives target FocusP or a more specific ExclamativeP in the focus field):

(9) [ExclamativeP *what an incredible week* [Exclamative′ EXCL [IP/TP *we have had*]]]

A logical extension of the account in (9) for cases of exclamatives with a low *that* (cf. (1)d) would be to hypothesize that the complementizer that is adjacent to the wh-phrase is the spellout of the Exclamative head since it co-occurs with the focal phrase:

(10) [ExclamativeP *what an incredible week* [Exclamative′ *that* [IP/TP *we have had*]]]

A similar account can be adopted for the other types of heads (topic, Q, etc.), as we will see below. This will indeed be the null hypothesis adopted in what follows. In the remainder of this paper, I will try to determine whether a Rizzian analysis is tenable for (all) the cases under discussion (cf. (1)–(2)) and whether a unified account can be proposed for all the cases at hand, a non-trivial issue given the complex dataset this paper concerns itself with. We will then consider this type of account more generally in terms of inter- and intra-linguistic variation in Section 4.

3. Different Left-Edge-Related Constructions: To Spell *That*/*Que* or Not to Spell *That*/*Que*

The paradigm in (1) and (2) attests to the complexity of the left periphery of languages like English and Spanish, where different constructions can occur—and even co-occur—in the left portion of the clause, and on occasion these constituents may be accompanied by an overt instance of *that/que*. Below, I concentrate on a subset of those constructions, for the reasons discussed in the previous section. I follow the relative order claimed by a

left-peripheral analysis like Rizzi's (cf. (7)). I first look at the English case and then at the Spanish one. Different existing accounts are considered, and mention is made of micro- (i.e., dialectal) variation when appropriate. In Section 4, I return to overarching existing accounts of the type of variation observed. I begin by discussing high complementizers.

3.1. High That/Que

High complementizers (by assumption, the head of C^{o} or Force^{o}, as in (4)) constitute an area of the grammar where languages like English and Spanish stand in glaring contrast to one another.

English, for its part, typically allows the complementizer to remain silent, as shown in (11), a feature that spreads across dialects:

(11) *I think Ø complementizers should not be taken for granted.*

The presence or absence of the overt complementizer in sentences like (11) does not have a semantic reflex, i.e., the truth-conditional meaning of the sentence remains intact. However, the presence of the complementizer is typically associated with academic or written registers. Certainly, using a complementizer facilitates reading and processing, and on occasion, it resolves potential ambiguities. This is shown by (12):

(12) *Johan said yesterday he accepted ten papers.*

In (12), *yesterday* can modify either s*aid* or *accepted*. Realizing the complementizer does away with this potential confusion:

(13) a. *Johan said that yesterday he accepted ten papers.*
b. *Johan said yesterday that he accepted ten papers.*

The optionality of *that* is not always such. For instance, topicalized clauses require *that* (Bošković and Lasnik 2003, among others):

(14) a. **Ø he accepted ten papers John believes.*
b. *That he accepted ten papers John believes.*

As for the subjunctive, conservative English speakers tend to disfavor the omission of *that*, contrary to what we observe in Spanish (cf. see the discussion surrounding (23) below) (Radford 2016), as in (15):

(15) *I demand that the children be here at ten.*

However, *that*-drop in subjunctives has also been documented in present-day English, even in writing, as shown by the below examples:

(16) a. *Brexit Secretary David Davis was also able to meet business leaders demanding Ø the UK stay in the single market immediately after leaving the EU* (*The Independent*, 10 July 2017, cited in Villa-García (2019, p. 26)).
b. *Following further evaluation this morning, The Queen's doctors are concerned for Her Majesty's health and have recommended Ø she remain under medical supervision* (statement from Buckingham Palace, 8 September 2022).
c. *Díaz evoked Dorado again on Monday, demanding Ø Feijóo explain how he knew the man* (politico.eu, 18 July 2023).

Needless to say, other well-known cases, such as comp-t(race) effects, force the complementizer to be absent in most dialects of English:

(17) a. *Who do you believe rocks?*
b. **Who do you believe that rocks?*

Furthermore, certain predicates have been claimed to require an obligatory *that* (Hegarty 1992; Adger 2003; Franks 2005; Radford 2016; Llinàs-Grau and Bel 2019, among others). The list includes verbs like *whisper*, *quip*, *judge*, and *conjecture*.[2] Nonetheless, a simple Google search yields examples featuring such verbs and no overt complementizer, as illustrated for *quip* in (18):

(18) *He correctly replied 14, then quipped Ø he wished it was 15* (*The Sun*, 2011, cited in www.collinsdictionary.com).

Early theoretical accounts of the overt/null contrast in complementizer lexicalization in languages like English in the Chomskyan tradition deemed the presence vs. absence of *that* a phonological phenomenon. More specifically, Chomsky and Lasnik (1977) advanced an analysis whereby *that* is deleted when absent (i.e., our null complementizer above would be the result of phonological deletion: ~~*that*~~). By contrast, Radford (2016) argues, on the grounds that the overt complementizer *that* and its null homolog *Ø* are not interchangeable in all contexts (e.g., (17)), that *that* and *Ø* are actually distinct lexical items exhibiting distinct properties (rather than a single item with two different spellouts, in the spirit of Chomsky and Lasnik (1977)).

According to Bošković and Lasnik (2003), null C (i.e., our *Ø* above) is a PF-affix that requires PF-adjacency with the verbal (or nominal) host. This requirement is fulfilled in (11) above, since the complement clause headed by C is PF adjacent to the V head, but not in (14)a, where the clausal complement has been fronted, which disrupts the necessary adjacency between the verbal head and C, rendering null C impossible and thus enforcing *that*, as in (14)b.

It is worth mentioning that a different line of analysis, adopted in the work of Pesetsky and Torrego (2001), takes *that*-less clauses to be the result of the subject moving to Spec, CP. In their view, *that* is the outcome of T^{o} moving to C^{o} for purposes of feature checking; this operation is satisfied if the subject moves to Spec, CP instead, resulting in *that*-less clauses. Put another way, for these authors, in *I think syntax is fascinating*, the subject *syntax* ends up in Spec, CP.

In stark contrast to what we observe for English, the high complementizer is generally mandatory in present-day Spanish, as shown by (19), a well-known cross-linguistic difference (Torrego 1983; Etxepare 1996; Brovetto 2002; Antonelli 2013; Cerrudo-Aguilar 2014; and Rodríguez-Riccelli 2018).

(19)	*Creo*	*que/*Ø*	*los*	*complementantes*	*no*	*se*	*pueden*	*dar*	*por sentados.*
	believe	that	the	complementizers	not	cl.	can	give	for seated

'I believe that complementizers cannot be taken for granted.'

On the other hand, diachronic and synchronic dialectal evidence suggests that the high complementizer is not obligatory at all times. For one thing, classical Spanish from the 17th century displayed null complementizers with certain predicates:

(20)	*Les*	*dijo*	*Ø*	*tenía nuevas*	*de*	*que*	*en*	*el*	*cielo*
	cl.	said		had news	of	that	in	the	heaven
	se	*había*		*muerto*	*el*	*arcángel*		*san*	*Gabriel*
	cl.	had		died	the	archangel		St.	Gabriel

'S/He said to them that s/he had news that the archangel St. Gabriel had died in heaven' (Abarca de Bolea, *Vigilia*, 17th century, cited in RAE-ASALE (2009, p. 3232)).

As far as diatopic variation at present is concerned, varieties that include Mexican Spanish have been reported to omit the complementizer with thinking and judgment predicates (Silva-Corvalán 1994 et seq.; Rodríguez-Riccelli 2018), as in (21).

(21)	*Creo*	*Ø*	*llamará.*
	believe		will-call

'I believe s/he will call' (RAE-ASALE 2009, p. 3232).

Using Tweeter data from Mexican and Los Angeles Spanish, Rodríguez-Riccelli (2018) concludes that verb modality and embedded subject position are the strongest predictors of *que*-drop in these varieties. Moreover, Rodríguez-Riccelli (2018) reports a higher percentage of *que*-drop in Mexico City (12.5%) than in Los Angeles (10%), suggesting that a contact-induced language change (from English) is not apparent. English contact, however, may allow the broadening of the contexts in which the omission of the complementizer occurs (e.g., stative verbs).

However, deletion can occur more generally across varieties under certain conditions. For instance, deletion of subordinating *que* occurs across Spanish varieties in formal-register contexts in cases where a mark of subordination (e.g., another *que*, an expression which includes *que*, such as *porque* 'because,' etc.) appears (Subirats-Rüggeberg 1987, pp. 170–71; Etxepare 1996; RAE-ASALE 2009):[3]

(22) *Recurrirán* *la* *sentencia,* *porque* *dicen Ø* *no*
contest the sentence because say not
es *ajustada.*
is fit
'They will contest the sentence, as they claim it is not appropriate.' (Telediario 2, RTVE, Spain, 16 April 2018)

In fact, across Spanish we find that subjunctive cases involving verbs such as *rogar* 'beg' tend to appear in formal and written contexts without *que* (the *que* counterpart of (23) being grammatical but less formal):

(23) *Rogamos* *Ø* *nos* *envíen* *el* *certificado* *a* *la* *mayor* *brevedad.*
beg cl. $\text{send}_{\text{Subjunctive}}$ the certificate at the bigger brevity
'We would like to ask you to send the certificate to us at your earliest convenience.'

At this juncture, two questions arise in light of data like (23):

(i) Are such cases instances of Force^{o} (high complementizers) or Finiteness^{o} (low complementizers), since they are related to mood (associated under Rizzi's proposal with FinitenessP)?

(ii) Is complementizer deletion the result of the verb moving all the way to C^{o} (Force^{o} or Finiteness^{o}, under Rizzi's assumptions)? (See, in this connection, the related claim noted above by Pesetsky and Torrego that *that*-less clauses in English arise from subject movement to CP).

A relevant question is also whether the non-appearance of *que* points to the absence of a left periphery altogether (so that such clauses are analyzed as bare IPs/TPs), a claim that would also extend in principle to English *that*-less clauses (Bošković 1997; Brovetto 2002; Antonelli 2013).[4] Antonelli (2013) argues that the left periphery of Spanish is present even in the absence of *que*, but that in cases like (23), the verb moves to a syncretic ForceP/FinitenessP projection, thus rendering the complementizer impossible. One problem with this type of account is that clitics show up preverbally (cl.+V—*les rogamos*, as in (23)), while in imperatives, which are standardly assumed to involve T^{o}-to-C^{o} movement, clitics show up postverbally (V+cl.—*ruégales* '$\text{beg}_{\text{imperative}}$ them'). The answers to (i) and (ii) await further research.

In sum, we observe that English and Spanish high complementizers behave quite differently, with the Spanish high-complementizer drop being much more restricted than its English counterpart.[5] More specifically, a (diachronic and synchronic) dialectal split emerges from our discussion surrounding Spanish: in Old Spanish, *que* could be dropped with verbs like *decir* 'say'; in present-day Spanish, *que* can only be absent in a very limited set of contexts (e.g., subjunctives with verbs like *rogar* 'beg'); in varieties such as Mexican or Los Angeles Spanish, *que*-less examples are confined to certain thinking and judgment predicates. Moreover, from an analytical perspective, the debate as to the analysis of complementizer-less sentences remains alive at present. We now turn our attention to what has widely been regarded so far as instantiations of non-high complementizers.

3.2. Recomplementation That/Que

The phenomenon of double-complementizer sentences, also known as recomplementation, illustrated again for English and Spanish in (24), has been subject to much debate in the recent literature (Escribano 1991; Iatridou and Kroch 1992; Demonte and Fernández-Soriano 2009; Villa-García 2012, 2015, 2019; Radford 2018; Villa-García and Ott 2022, amongst many others).

(24) a. *Note that if you have already taught in Semester 1, that you are not required to resubmit paperwork to HR Services* (official university communication, UK, January 2019, cited in Villa-García (2019, p. 2)).

b.

Dice	*que*	*si*	*llueve,*	*que*	*se*	*quedan*	*encamados*
says	that	if	rains	that	cl.	stay	bedded

'S/He says that if it rains, that they will stay in bed.'

As regards syntactic microcomparison in the realm of recomplementation, no noteworthy differences are reported in the existing works in relation to English; the data provided in the literature come from both American and British English (Radford 2018; Villa-García 2019), but little or no attention has been paid to whether there exists dialectal variation in English recomplementation, a gap in the literature at present.

As for Spanish, whereas Demonte and Fernández-Soriano (2009) note that recomplementation is found across Spanish varieties, Martínez-Vera (2019) claims that recomplementation is absent in American Spanish (its counterpart being a prolonged intonational break); however, Frank (2020) provides experimental evidence from Colombian and bilingual heritage US Spanish indicating that recomplementation is not impossible in such varieties. Similarly, linguists such as Andrés Saab and Carlos Echeverría (pers. comm. 2022), who are speakers of Argentinean and Chilean Spanish, respectively, use and accept recomplementized structures. Fontana (1993) and Echeverría and López Seoane (2019), for their part, observe, based on written evidence, that Old Spanish frequently featured recomplementation, the sandwiched elements being typically long *if*-clauses. Needless to say, the foregoing discussion strongly suggests that the dialectal map of recomplementation in Spanish is likewise in dire need of further research.

Be that as it may, early proposals (e.g., Iatridou and Kroch 1992) assumed that the different *that* complementizers featured in what looks like a single sentence whose embedded clause displays a complex left periphery are instances of C^{o} in a recursive CP, as in (25) (see also (6) above):

(25) $[_{CP}$ $[_{C'}$ *that/que* $[_{CP}$ XP $[_{C'}$ *that/que*]]]]

In the wake of the Rizzian approach, the question soon arose as to which head is spelled out by doubled, secondary complementizers. A myriad of proposals arose (on which see Villa-García (2015)); I will concentrate on two here for the sake of illustration. Authors like Demonte and Fernández-Soriano (2009), López (2009), and Antonelli (2013) have advanced the hypothesis that the high and the low complementizers delimit the beginning and the end of the left-peripheral space, hence populating Force^{o} and Finiteness^{o}, respectively:

(26) $[_{ForceP}$ $[_{Force'}$ *that/que* $[_{TopicP}$ XP $[_{Finiteness'}$ *that/que*]]]]

However, such an analysis runs into a number of empirical problems, as argued by Villa-García (2012, 2015, 2019). Instead, this author proposes to treat the second instance of *that/que* as a topic marker, hence the head of TopicP (see also Rodríguez-Ramalle (2003); see Ledgeway (2005) for an analysis of this type which assumes that the different complementizers are separate realizations of a complementizer that starts in Finiteness^{o} and moves up to Force^{o} in a head-to-head fashion):

(27) $[_{ForceP}$ $[_{Force'}$ *that/que* $[_{TopicP}$ XP $[_{Topic'}$ *that/que* $[_{FinitenessP}$ $[_{Finiteness'}$]]]]]]

This analysis is *prima facie* appealing both on empirical and theoretical grounds (it accounts for why it is typically topical phrases that appear in a sandwiched position, and it assumes that the different left-peripheral heads proposed by Rizzi (1997) can be spelled out).

Nevertheless, in marked contrast to monoclausal proposals like the ones just reviewed, recent research has convincingly argued that recomplementation is not bound to occur only with topical phrases and that in fact what doubled complementizers mask is two separate sentences that superficially look like one, rather than an elaborate clausal left edge (see, especially, Villa-García and Ott (2022) on this biclausal line of analysis), thus arguing against monoclausal accounts like those in (26) and (27). For these authors, the clausal

portion headed by the second instance of *that/que* represents a restart in discourse (i.e., a new sentence, CP2, reprising the first one, CP1, begins), as shown abstractly in (28):

(28) [$_{CP1}$ subject V [*that/que* … XP]] [$_{CP2}$ ~~subject V~~ [*that/que* …]]

This move accounts for issues including why the second sentence must be a syntactically complete sentence (i.e., *[that he/*Ø hates seafood]* in (29)a), why a non-topical phrase (e.g., a discourse marker, as in (29)b), can occur in between overt complementizers, and why even a focal phrase can appear in between *that*s/*que*s (cf. (29)c):

(29) a. *He told me that Peter, that he/*Ø hates seafood.*
b. *Dice que bueno, que no vienen.*
says that well that not come
'S/He says that well, that they are not coming.'
c. *Me dijo que jamás, que jamás se casará con nadie.*
cl. says that never that never cl. will-marry with nobody
'S/He says that never ever, that never will s/he marry anyone.'

For Villa-García and Ott (2022), therefore, the second instance of the complementizer is a restart that mirrors what we see in the first clause (i.e., *dice que/says that*… — ~~*dice que/says that*~~…), hence a repeated high complementizer, but not the realization of a left-peripheral head such as Topic^{o} (or Finiteness^{o}, for that matter), as assumed in (26) and (27). This actually goes a long way to explaining why we also find reduplicative complementizers other than declarative *that/que* (i.e., *he asked me whether*…— ~~*he asked me*~~ *whether*…):

(30) *He asked me whether, given the current assessment boycott, whether we are getting a salary raise.*

In light of data like (29) and (30), the prospect that recomplementation *que* is the overt or null spellout of Topic^{o} under monoclausality (*that/que* vs. ~~*that/que*~~) loses plausibility, which casts doubt on the claim that Topic^{o} can be occupied by an overt realization of the complementizer *that/que*. On the bisentential account, therefore, recomplementation *that/que* would be an instance of high *that/que* in disguise. Technically, then, *that/que* is the same high element in the two occurrences (in CP1 and CP2), in spite of outward appearances. In the next two subsections, I turn to the exclamative and interrogative *that/que*.

3.3. Exclamative That/Que

Exclamative wh-phrases that come in the company of *that* have customarily been attributed to Irish English (Zwicky 2002). However, Radford (2018) provides a large set of data suggesting that other varieties permit the co-occurrence of wh-phrases with an instance of *that* below them as well, as shown again in (31):

(31) a. *How gorgeous that you look!* (www.abbieeandeveline.com, cited in Radford (2018, p. 160)).
b. *How quickly that people forget!* (web, cited in Radford (2018, p. 159)).
c. *What a job that he's done so far!* (Sam Matterface, Talksport Radio, cited in Radford (2018, p. 159)).

Of course, the *that*-less counterparts of the examples in (31) would be the canonical versions of the relevant sentences, which shows that *that* is once again not mandatory. Radford (2018) pursues an account in the spirit of Rizzi (1997) and Rizzi and Bocci (2017) wherein the wh-phrase in the specifier of ExclamativeP (or FocusP) licenses the head *that*:

(32) [$_{ForceP}$ [$_{Force'}$ [$_{ExclamativeP}$ XP [$_{Exclamative'}$ *that*]]]]

An issue that any analysis needs to tackle concerns the rare occurrence of wh-items (or wh-words) in this context. For the most part, it is almost always a full wh-phrase that appears immediately above *that* in exclamatives in English. Authors like Bayer (2014) and Radford (2018) have argued that unlike wh-phrases, which move to the specifier, wh-words move to the head position of the wh-operator projection (e.g., FocusP), thus preventing a complementizer from occurring in such a position (i.e., wh-phrase + *that* vs. *wh-word + *that*).

Structures like those in (31) call into question longstanding claims in the generative literature, including the Doubly-Filled Comp Filter, which bans the simultaneous occurrence of a wh-element and an overt complementizer in CP, a prohibition observed in standard English (i.e., *[$_{CP}$ wh-element [$_{C'}$ *that* . . .]]). In his discussion of similar examples of exclamative + *che* 'that' examples in Italian, Rizzi (2013, p. 2009) actually refers to this configuration as "the only case of legitimate '[D]oubly[-]filled Comp' in Standard Italian" (see also Bayer and Dasgupta (2016)). Rizzi goes on to note that "[c]learly, (the equivalent of) *that* is an unmarked, versatile complementizer form, capable of occurring in the highest C position, and also, in cross-linguistically variable manners, in lower positions", as in (32).

Moving from root to embedded contexts, Radford (2018) provides data indicating that it is not impossible to have a wh-exclamative below an instance of what appears to be a high (Force^{o}) complementizer (a well-known property of Spanish, as we shall see):

(33) *He realized, I think, that how big this thing was* (Film critic, BBC Radio 5, cited in Radford (2018, p. 114)).

Radford (2018) does not give any examples of embedded wh-exclamatives plus *that*, but such examples sound fine to his native ear:

(34) *He realized that how big that this thing was* (Andrew Radford, pers. comm. 2023).

Spanish exclamatives behave similarly to their English counterparts reviewed above in most respects, with interesting dialectal differences. In many varieties of Spanish, a pleonastic complementizer immediately adjacent to the wh-phrase is not unusual (speakers notice that the *que* version is more emphatic):

(35) a. ¡ *Qué* *guapa* *que* *está* *tu* *niña!*
what beautiful that is your daughter
'Your daughter looks so good!'

b. ¡ *Qué* *rápido* *que* *conducen* *aquí!*
what fast that drive here
'They drive so fast here!'

A notable difference with present-day English is that exclamatives in (non-Caribbean) Spanish trigger obligatory S-V inversion regardless of the presence vs. absence of *que* (see Villa-García (2018, in preparation) and Villalba (2019) for recent discussion; as noted by an anonymous reviewer, Italo-Romance varieties follow a similar pattern):[6]

(36) *¡ *Qué* *guapa* *que* *tu* *niña* *está!*
what beautiful that your daughter is
Intended: 'Your daughter looks so good!'

Certain dialects (e.g., Asturian Spanish) disallow the presence of *que* with *qué* phrases (Villa-García 2018, in preparation). By contrast, the popular Spanish of Asturias allows *que* with the exclamative determiner *vaya* (on *vaya* more generally, see, e.g., Espinal et al. (2022)). According to Bosque (2017, pp. 18–19), *vaya* in Asturian Spanish (and in areas of León) can combine not just with nouns (which is what happens in other parts of the Spanish-speaking world), but also with adjectives and adverbs:

(37) a. ¡ *Vaya* *casa* *que* *tiene!*
what house that have
'What a house s/he has!'

b. ¡ *Vaya* *sano* *que* *está!*
what healthy that is
'How healthy he is!'

c. ¡ *Vaya* *mal* *que* *lo* *hiciste!*
what bad that cl. did
'You did it so badly!'

In Spanish, cases like (37)a involving nouns can optionally have *que* (RAE-ASALE 2009). Asturian Spanish, though, exhibits a more nuanced contrast with respect to *vaya* exclama-

tives, irrespective of the category of the word following *vaya*. In areas such as Avilés and Gijón, the *que* version is the norm, with the non-*que* version being deemed incomplete or even unacceptable. However, in Oviedo and the surrounding hamlets, the *que*-less version prevails (perhaps due to influence from the Asturian language). To summarize, in Asturias, *qué* + phrase + *que* is not an option; only the *que*-less version is used. When it comes to *vaya* + N/A/Adv, however, we find subdialects that require *que* and others that dispense with it. In other words, the form of the exclamative particle determines the possibility of having an accompanying *que* below the exclamative phrase in these varieties.

The evidence furthermore shows that there is a dependence between the presence vs. absence of *que* and the type of phrase in the specifier (Villa-García, in preparation): *que* is sensitive to the sort of exclamative element to its left, which confirms that the two stand in a spec-head configuration, exactly as predicted by accounts like (32) above, which invoke FocusP/ExclamativeP. This analysis is substantiated by the obligatory inversion displayed by exclamatives, which is characteristic of focal phrases in (non-Caribbean) Spanish (on this issue, however, see Villa-García, in preparation). Thus, I will adopt the account in (32) for Spanish as well, as in (38):[7]

(38) $[_{\text{ForceP}}$ $[_{\text{Force}'}$ $[_{\text{ExclamativeP}}$ XP $[_{\text{Exclamative}'}$ *que*]]]]

Beyond Peninsular Spanish, according to RAE-ASALE (2009, p. 3206), in Latin American Spanish, the presence of *que* with *qué* exclamatives is more restricted than in Spain, although examples occur in the River Plate area and less frequently in the Caribbean (see also Casas (2004, p. 268) for examples from Mexican Spanish).

As far as embedded clauses are concerned, pleonastic *que* is also acceptable in non-matrix contexts:[8]

(39) *Mira* *qué* *guapo* *que* *es* *ese* *podcáster*.
look what good-looking that is that podcaster
'Look at how good-looking that podcaster is' (Antonio Cañas García, Raquel González Rodríguez, and Isabel Pérez-Jiménez, pers. comm. 2023).

Still, exclamative clauses selected by predicates other than pseudo-interjections like *mira* 'look!' cannot be construed with *que*, a poorly understood phenomenon to date (Ignacio Bosque, pers. comm. 2023):

(40) *Es* *increíble* *qué* *cosas* *(*que)* *dice*.
is incredible what things that says
Intended: 'The things s/he says are incredible' (Bosque 1984, p. 287).

Villa-García (2015) furnishes embedded data with *vaya* under verbs of saying in Asturian Spanish, along the lines of (41):

(41) *Dice* *mi* *prima* *que* *vaya* *rápido* *que* *conduce* *tu* *padre*.
says my cousin that what fast that drives your father
'My cousin says that your father drives so fast.'

Qué-exclamatives plus *que* are also licit under *decir* 'say'-like predicates. Note that English exhibits this pattern as well (cf. (33) and (34)):

(42) *Dice* *que* *qué* *guapa* *que* *es* *esa* *niña.*
says that what beautiful that is that girl
'S/He exclaimed that that girl is so beautiful' (Raquel González Rodríguez and Isabel Pérez-Jiménez, pers. comm. 2023).

All in all, the English and Spanish evidence adduced here points to the conclusion that the exclamative phrase and *que* co-exist in the same projection. Nonetheless, dialect data from Spanish point out that the exclamative *que* is not really optional at all times. While in many areas of Spain exclamatives with *qué* and *vaya* seem to optionally co-occur with the pleonastic *que*, *qué*-exclamatives in Asturian Spanish occur without *que*; their *vaya* homologs require *que* in some parts of Asturias, but not in others. The data crucially corroborate that the licensing of *que* is sensitive to the nature of the wh-phrase in its specifier, which supports

a spec-head analysis along the lines of (38). This account indeed gains cross-linguistic plausibility from the English data discussed above, which are not confined to Irish varieties of English, as used to be widely thought. Lastly, the evidence also reveals a high degree of variation across Spanish varieties, which further studies should certainly investigate in more depth. The facts are also relevant to non-trivial theoretical questions, including whether the cases at hand are compatible with a relative-clause analysis (on which see fn. 7). I now turn to wh-interrogatives with *that/que*.

3.4. Interrogative That/Que

Wh-interrogatives followed by *that* used to be believed to be a feature of regional varieties like Irish English, as in the Belfast English example in (43) (with some Belfast speakers accepting only wh-phrases –not wh-items– above *that*, much like in the wh-exclamative cases discussed above):

(43) *I don't know when that he's going* (Henry 1995, p. 88).

Radford (2018) shows, by contrast, that wh-interrogative + *that* configurations transcend Irish varieties, as the data in (44) demonstrate:

(44) a. *Definitions vary as to which of these types of criteria that are used* (Member of the English Department, University of Göteborg, cited in Radford (2018, p. 137)).
b. *I hadn't realized just how many people that were there*
(Maxx Faulkner on WCBE, cited in Radford (2018, p. 138)).
c. *This heat map shows just how active that Trippier was*
(Jermaine Jenas, BBC1 TV, cited in Radford (2018, p. 139)).

Radford (2018, p. 142) submits that the above data are amenable to a Rizzian account according to which the wh-interrogative is housed in the specifier of a WHP (or FocusP) below ForceP, with *that* in the head position of WHP/FocusP:

(45) $[_{\text{ForceP}}$ $[_{\text{Force}'}$ $[_{\text{WHP}}$ wh-interrogative $[_{\text{WH}'}$ *that* . . .]]]]

One advantage of this account is that it can easily accommodate cases of embedded wh-interrogatives below a quotative element, reported in Radford (2018, p. 116), and which are used "to embed quoted speech into a matrix clause, with the quoted speech essentially being unmodified", as in (46).

(46) a. *He protested that how could he have known that his office was bugged?* (Radford 2018, p. 113).
b. $[_{\text{ForceP}}$ $[_{\text{Force}'}$ *that* $[_{\text{WHP}}$ wh-interrogative $[_{\text{WH}'}$ . . .]]]]

With verbs that intrinsically select a question as their complement, the secondary *that* is legitimate, but not the high *that*, as the below example, kindly provided by Andrew Radford (pers. comm. 2023), illustrates. This is a non-trivial dissimilarity between English and Spanish, where embedded interrogatives can be heralded by an instance of reportative *que* preceding either interrogative phrases or the interrogative complementizer (on which see Plann (1982) and much subsequent work; cf. (48)b).

(47) *I wonder (*that) what kind of party (that) he has in mind.*

As far as Spanish is concerned, the literature notes an important paradigm gap owing to the non-existence of the low *que* with wh-interrogatives either in root or in embedded contexts, in sharp contrast to English and to what we observe in the case of Spanish exclamatives in the preceding subsection:

(48) a. *¿ Cuántas casas (* que) se ha comprado?*
how-many houses that cl. has purchased
'How many houses has s/he bought?'
b. *Preguntaron (que) cuántos kilómetros (* que) había recorrido.*
asked that how-many kilometers that had travelled
'They asked how many kilometers I/s/he had travelled.'

The lack of *que* with wh-interrogatives in Spanish is widely considered to be the only formal mark distinguishing interrogatives from exclamatives as far as their structure is concerned (RAE-ASALE 2009, among others).[9] As we shall see momentarily, however, there are dialectal data indicating that, on occasion, *que* surfaces in certain interrogative contexts, which refutes the widely held generalization that *que* never co-occurs with wh-interrogatives in Spanish.

Before we delve into the dialectal data in question, it is important to note there is a well-documented interrogative phrase that occurs with *que* in all dialects of Spanish, however. This is the *cómo que* 'how come' (lit. 'how that') construction. But unlike regular interrogatives, this one is a formulaic construction that triggers no inversion:

(49) *¿ Cómo que al final no vienes al bodorrio?*
how that at+the end not come to+the wedding
'How come you are finally not coming to the wedding party?'

Importantly, some varieties of Latin American Spanish permit certain interrogatives to occur with a low instance of *que*, *contra* standard contentions regarding the impossibility of having *que* immediately after a wh-phrase across Spanish, as in (48). The following data from CORPES XXI, generously furnished by an anonymous reviewer, show that this is in fact the case (see also Villa-García, in preparation):

(50) a. *¿ Desde cuándo que no lo ven?*
since when that not cl. see
'When did you last see it/him?' (Chile).

b. *¿ Cuándo fue y dónde que ocurrió ese descubrimiento?*
when was and where that occurred that discovery
'When was it and where was that discovery made?' (Uruguay).

c. *¿ De dónde que alguna vez en otra vida lejana, había pretendido y creído ser escritor?*
of where that sometime in other life far had intended and believed be writer
'Where did you learn that, some other time, in a different, distant life, he had intended to be and believed himself to be a writer?' (Cuba).

d. *¿ Por qué que no fuiste a rescatarnos?*
x for what that not went to rescue-cl.
'Why did you not go to rescue us?' (Colombia).

e. *¿ Y por qué que no me arriesgaría a algo así?*
and for what that not cl. risk to something thus
'And why wouldn't I risk doing something like that?' (Chile).

The wh-phrase-plus-*que* data just reviewed raise various questions, such as whether this pattern can be found with other wh-items or is confined to adjuncts, whether the interpretation of the sentences displaying *que* is different from that of their *que*-less homologs, and whether they can occur in subordinate environments. This pattern is actually well documented in other Romance languages like Brazilian Portuguese, Canadian French, and the Northern Italian dialect of Lamonat (Simone De Cia, pers. comm. 2023), as shown by the following example from Brazilian Portuguese:

(51) *Onde que você mora?*
where that you live
'Where do you live?' (Oushiro 2011, p. 145).

Overall, we do find a subset of varieties of Spanish where the wh-interrogative + *que* configuration is legitimate. An analysis along the lines of that for English can therefore be advanced for these cases, as in (52). The question of course still remains as to why the presence of *que* below interrogatives in Spanish is so highly restricted.

(52) $[_{\text{ForceP}}$ $[_{\text{Force}'}$ $[_{\text{WHP}}$ wh-interrogative $[_{\text{WH}'}$ *que*]]]]

Having discussed exclamatives and interrogatives, we now turn our attention to what has been assumed to be the lowest complementizer along the leftmost part of clauses: 'jussive/optative' *that/que*.

3.5. Low Complementizers: 'Jussive/Optative' That/Que

A final phenomenon that I will consider here is the so-called 'jussive/optative' complementizer (Villa-García 2012, 2015; Radford 2018), which I touched upon in passing in Section 3.1. This phenomenon is illustrated again for English below:

(53) *I am writing to ask* ***that*** *if you have not yet completed this training in this academic year* ***that*** *you do so as soon as possible and by the end of 14 July 2023 at the latest* (Official university communication, United Kingdom, 30 June 2023).

That in these cases is deemed to be the lexical realization of the subjunctive. As noted, conservative speakers do not drop *that* in this context, as illustrated once more in (54), but present-day English allows *that*-omission (cf. (16)).

(54) *The University has ordered that a town be built in the premises.*

The question which arises is whether 'jussive/optative' cases lexicalize Force^{o} or rather Finiteness^{o}, which is connected to mood under Rizzi's (1997, et seq.) system. One possibility, entertained by authors like Rizzi (1997), Villa-García (2012, 2015), Antonelli (2013), and Radford (2018), is that in the absence of sandwiched material, a conflated ForceFinitenessP is projected. In this context, we are no longer dealing with high or low *that*, since a unique realization would do the job under ForceFiniteness^{o} (which is, in effect, equivalent to a C^{o} analysis like that in (4)):

(55) [$_{\text{ForceFinitenessP}}$ [$_{\text{ForceFiniteness}'}$ *that*]]

By contrast, when left-peripheral constituents occur, which is when multiple instances of *que* surface (i.e., *that* ... XP ... *that*), a split of the CP field is triggered (Rizzi 1997; Villa-García 2015; Radford 2018, among many others). Analytically, it is conceivable that this instance of low *that* may be a lexicalization of Finiteness^{o}, as follows:

(56) [$_{\text{ForceP}}$ [$_{\text{Force}'}$ *that* [$_{\text{TopicP}}$ XP [$_{\text{Topic}'}$ [$_{\text{FinitenessP}}$ [$_{\text{Finiteness}'}$ *that*]]]]]]

This proposal receives empirical support from other linguistic varieties. For instance, Ledgeway (2005, p. 365) capitalizes on languages like Romanian, which exhibits a distinct (low) complementizer (*să*) in subjunctive clauses that appears to be a very low element in the left-peripheral spine, as witnessed by its mandatory proximity to the verb and any clitics that may come with the verb (that is, any left-peripheral phrase must precede *să*). Romanian (57) illustrates (as observed by an anonymous reviewer, other Balkan languages make the same point):

(57) *Vreau (ca) MÂINE să meargă.*
want that tomorrow that go$_{\text{Subjunctive}}$
'I want him to go TOMORROW' (Watanabe 1996, p. 44).

Such examples are ideal candidates for an analysis in the spirit of (56) above:

(58) [$_{\text{ForceP}}$ [$_{\text{Force}'}$ *ca* [$_{\text{FocusP}}$ *MÂINE* [$_{\text{Focus}'}$ [$_{\text{FinitenessP}}$ [$_{\text{Finiteness}'}$ *să*]]]]]]

Villa-García (2012, 2015, 2019) has made the contention that Spanish also possesses a Finiteness^{o} complementizer lexicalized as *que*, which he dubs 'jussive/optative' *que*, exemplified by the following data:

(59) a. *Juan Carlos también le dijo a su hijo que si tuvo el coraje de casarse con ella*
Juan Carlos also cl. told dat his son that if had the courage of marry-cl with her
desobedeciendo el deseo de sus padres que tuviera el carácter
disobeying the wish of his parents that had$_{\text{Subjunctive}}$ the temper
para ponerla en su sitio.
for put-cl. in her place
'King Juan Carlos I also demanded that if his son had the courage to marry her, despite his parents' wishes, that he muster the strength to put her in her place.'
(paraphrase of Jaime Peñafiel's words, *Lecturas*, www.lecturas.com, 8 April 2018)

b. *Que si vas a salir con ella, que vayas en serio.*
that if go to go-out with her that go$_{\text{Subjunctive}}$ in serious
'I'm saying that if you are going out with her, you should get serious' (RTVE, Servir y proteger, TV series, 4 April 2018).

Villa-García (2018) reports emphatic examples from naturalistic speech where indeed the two *que*s actually co-occur (which can arguably be analyzed as simultaneous realizations of Force^{o} and Finiteness^{o} under Rizzi's system):

(60) a. *Que te ha dicho que que te pires.*
that cl. has said that that cl. go$_{\text{Subjunctive}}$
'I'm telling you s/he told you to go away.'

b. *Así que que te den.*
so that that cl. give$_{\text{Subjunctive}}$
'So go fuck yourself.'

An immediate question begged by the subjunctive data reviewed so far is whether this instance of low *que* is compulsorily overt or not, vis-à-vis recomplementation *que* in Section 3.2, which is optional (see Villa-García (2015) and Echeverría (2020) for much relevant discussion). A preliminary survey suggests that speakers prefer the realization of this low *que*, but not all speakers fully reject the *que*-less counterparts (a claim that can be extended to apply to the English cases above). Thus, when it comes to 'jussive/optative' cases in subordinate contexts, we are dealing with a matter of preference, rather than obligatoriness. Echeverría (2020, p. 48, fn. 24) furnishes the following example, which he attributes to the Royal Academy of the Spanish Language, indicating that 'jussive/optative' *que* can be silent:[10]

(61) *Dile que cuando termine venga a rendir cuentas.*
say that when finishes come$_{\text{Subjunctive}}$ to accounts accounts
'Tell him/her to come (here) to give us an explanation when s/he's done.'

There may be factors such as tense (present vs. past), the presence of negation, intrinsically subjunctive selecting predicates (*pedir* 'request') vs. communication verbs (*decir* 'say'), or even diatopic variation at play here. In fact, speakers of Spanish in contact with Catalan seem more permissive in terms of low *que*-omission (Villa-García 2018). Echeverría (2020, p. 48, fn. 24) arrives at the conclusion that "[i]f Spanish optative and jussive sentences are overall more likely to include an extra complementizer, this might well be explained by the more general, historically increasing tendency to use *que* before verbs in the subjunctive". Actually, this type of *que* also occurs in root clauses (Villa-García 2015, amongst others), in which case *que* is unquestionably compulsory:

(62) *¡Que venga a verme tu hija!*
that come$_{\text{Subjunctive}}$ to see-cl. your daughter
'I demand that your daughter come to see me.'

Returning to the embedded cases in (59) displaying a doubled *que*, whether obligatory or not, the evidence is symptomatic that there exists a low subjunctive complementizer in languages like Spanish.

Nevertheless, an open question not addressed by Villa-García and Ott (2022) is whether 'jussive/optative' sentences in English and Spanish can also be reanalyzed as restarts, much like their recomplementation homologs. An analysis of this guise for the cases at stake would assume the following preliminary structure for subordinate 'jussive/optative' sentences:

(63) [$_{\text{CP1}}$ subject V [*that/que* . . . XP]] [$_{\text{CP2}}$ ~~subject V~~ [*that/que* $V_{\text{subjunctive}}$]]

This move would easily account for why the subjunctive *que* is more likely to be realized in this environment: the second occurrence would be a repeat of the same element in a restart configuration, with complementizers heralding subjunctive clauses being less

omittable, as has been noted (see Section 3.2 for more details of this analysis when applied to recomplementation *que*). If this analysis ends up being the right account of reduplicative cases featuring "jussive/optative" *that/que* as well, then what I have referred to as medial (recomplementation) and low complementizers ('jussive/optative' *que*) would be underlying instantiations of a high *que* (see above on whether we are dealing with Force°, Finiteness° or even a conflated ForceFiniteness° projection when no overt left-peripheral material occurs).

Having discussed a major subset of the putatively different positions in which complementizers can be realized in English and in Spanish, the following section explores the more general and crucial issue of parameterizing the various lexicalization possibilities observed hitherto.

4. Inter- and Intra-Linguistic Variation: (Micro-)Parameterizing the Presence vs. Absence of Complementizers

The preceding discussion has made it clear that the seemingly ubiquitous *that* and *que* complementizers in English and Spanish constitute a gold mine for the study of the geometry of the clausal left edge as well as for variation, including macro- and micro-variation. In the following subsections, I turn to potential accounts of the general variation observed, based on recent proposals in the generative tradition. I will divide the discussion into spell-out accounts and feature-driven accounts, with a final note on a potential consideration regarding the detectability of projections that may help explain the different realizations found. A more general question raised by the discussion so far to which we will return towards the end of the paper is in fact whether a unitary account is attainable.

4.1. PF Accounts

The work of Rizzi and his collaborators has advocated a Spell-Out (i.e., pronunciation) Parameter analysis to account for the (non)-overtness of left-peripheral heads. In the words of Rizzi (2013, pp. 201–2), we are dealing with "a spell-out parameter, a familiar and widely attested kind of low[-]level parameterization". According to Rizzi and Bocci (2017, p. 13), the different left-peripheral criterial heads (Force°, Topic°, etc.) may be "null. . ., but their presence may be detected indirectly". The authors go on to say that ". . .the same relevant criterial [CP] heads are phonetically null, a familiar (and trivial) parametric difference". I propose that this type of parameter could be extended to apply across varieties of the same language, effectively making it a micro-parameter (see also Villa-García and Rodríguez (2020) on *sí (que)* 'yes that', exemplified in (2)c, across Spanish). An abstract visual representation of this binary parameter would be as follows, where ~~strikethrough~~ indicates PF deletion (note that the bracketings in (64) are highly simplified by only focusing on those projections that this work has specifically concentrated upon):

(64)
a. [ForceP [Force′ *that/que* [TopicP [Topic′ *that/que* [FocusP [Focus′ *that/que* [FinitenessP [Finiteness′ *that/que* IP/TP]]]]]]]
b. [ForceP [Force′ ~~*that/que*~~ [TopicP [Topic′ ~~*that/que*~~ [FocusP [Focus′ ~~*that/que*~~ [FinitenessP [Finiteness′ ~~*that/que*~~ IP/TP]]]]]]]

Analogously, on the basis of English data akin to those in (1) in the introduction, Radford (2018, p. 170) puts forth the following conditions for complementizer realization in English (note that % signposts inter-speaker variation):

(65) ***Complementizer spellout conditions*** (English)
In a finite clause, a non-verbal peripheral head can be spelled out as *that*
(i) if it is the first word in an embedded clause
or (ii) if it (% is in an embedded clause and) has an adjacent superordinate (% edgemate) (% non-wh) licenser

It is of note that this author uses the modal *can*, which reflects the optionality of *that* in English across the board, although it is important to recall that in some contexts, high *that* must be pronounced (Bošković and Lasnik 2003, *inter alia*; see Section 3.1).

What the pronunciation accounts outlined above share is the assumption that left-peripheral heads (for us, the complementizers *that/que*) are always optional, their realization boiling down to a superficial parameter that decides whether the relevant head positions are PF realized (i.e., lexicalized) or not. In spite of the fact that such accounts are a priori

theoretically appealing owing to their elegance and simplicity, they fall short of accounting for the facts in a satisfactory manner. For one thing, certain *that* realizations are more likely than others are (e.g., subjunctive *that* for conservative speakers) or even compulsory in some contexts (initial *that* in clause-fronting cases like (14)b). Similarly, once Spanish is brought into the picture, the wide range of variation brought to light herein demonstrates that limiting all variation to a mere (discretionary) pronunciation decision appears to be, at best, oversimplistic.

A broader issue concerns the conception of parameters as binary choices (Chomsky 1981). If truly a parameter, then [±spellout] should be set to one value (e.g., [+spellout] or [-spellout]) for a specific construction in one particular language/dialect, not to either option (i.e., a particular CP-related head would bear a specific value for the relevant feature in charge of parametric variation in this regard, but not varying values/settings; see below on the prospect of lexical parameters). Put another way, optionality should in principle be excluded, contrary to fact. Assuming that *that/que* may be null/overt when optional based on a parametric difference would be akin to saying that a given language can have the positive and the negative setting of the Null-Subject Parameter at the same time, on an optional basis. Unless ancillary stipulations are made, it is not at all clear how the Spell-Out Parameter option would work in practice for the cases at issue.

4.2. Feature-Based Accounts

In the wake of the Minimalist Program (Chomsky 1995), features acquired a prominent role in syntactic theory. Early works already hinted at the possibility of explaining differences in terms of complementizer lexicalization by means of features. For instance, in an attempt to rule out wh-interrogatives plus *que* in languages like Spanish (see Section 3.4), Uriagereka (1995, p. 160) observed that "although *que* can occupy the head…, in these languages [i.e., in Romance varieties such as Galician and Spanish, JVG] it does not have the appropriate features to agree with a focused phrase in its spec—much like *that* is not compatible with Wh-phrases [in standard English, JVG]" (see also Brucart (1993, p. 76) on a similar claim regarding exclamatives).

This is consonant with the prevalent idea that parametric variation is connected to features of functional heads (cf. the Borer-Chomsky conjecture; "lexical parameters"). In other words, this type of approach locates the relevant dimensions of (micro)variation in the properties of individual functional heads. As Ledgeway (2020, p. 31) puts it, "the locus of parametric variation lies in the lexicon, and in particular, in the (PF-)lexicalization of specific formal feature values of individual functional heads". The Minimalist Program, therefore, paved the way to recast parametric variation as different featural configurations encoded in lexical heads, in such a way that distinct featural make-ups yield the differences noted.

With particular reference to microvariation, a number of studies (e.g., Smith and Adger 2005 et seq.; Thoms et al. 2019) have shown that the Minimalist Program is particularly well suited to deal with such micro-variation (see Green (2007) for an overview of approaches to syntactic (micro-)variation). For example, Smith and Adger (2005) note that the feature system allows for variable phonetic (i.e., audible) outputs with the same interpretation (by virtue of a particular head containing one additional formal feature, with visible phonetic/PF consequences but not semantic/LF ones). This move clearly accounts for those cases where an element can be present or absent without meaning consequences. Critically, the feature-driven approach can also capture those cases where the meaning may change, such as exclamatives with *que* in Spanish, where most speakers report added emphasis in the presence of the overt *que*. Under this feature-based-type of analysis, such arguably semantic differences would be owing to different semantic features in the relevant head. Thus, my proposal would be that all of the C-related heads share some feature (possibly C), which is responsible for the phonological uniformity of the realization (*that/que*), but that various dialects allow deletion/non-realization depending on the other features of the head, in the spirit of Smith and Adger (2005).

4.3. Projection Detectability

Beyond spellout conditions and features, Spanish complementizer realization may show sensitivity to whether a projection has or lacks an overt specifier (with ForceP and FinitenessP being excellent candidates for the latter type). It seems that in those projections lacking a specifier, the head is generally less omittable across dialects (i.e., high and low *que*), in keeping with the generalization that languages disallow XPs to be headed by silent heads and specs (Koopman 1997) or, more generally, that a phrase is only projected when overtly manifested (Roberts and Roussou 2003; An 2007; Bošković 2016, *inter alia*). By contrast, when a specifier (e.g., Topic, Exclamative) is present (in apparent violation of the Doubly-Comp Filled Filter, as noted above in passing), the head can more easily remain null across dialects. The co-occurrence of an overt specifier with an overt head then leads to a reinforced/even more emphatic construction (cf. the emphasis contrast in *qué*-exclamatives with and without *que*, with *que* adding emphasis according to my native-speaker consultants).

This move goes against a superficial pronunciation or spell-out parameter (see Section 4.1); nevertheless, it is also challenged by a number of non-trivial questions: why is *que* obligatory with *vaya* for a subset of Asturian Spanish speakers (see Section 3.3)? What about the English case, where spec-less Force^{o} can remain silent (see Section 3.1)?[11] These and other questions strongly indicate that the ultimate account of the (non-)realization of left-peripheral heads in the form of *that/que* awaits and merits further investigation.

5. Conclusions

Since Rizzi's (1997) split-CP hypothesis, the investigation of the mapping of the left periphery has commanded the attention of numerous researchers working in a vast number of languages. Much care has been taken to study the different constructions (topics, exclamatives, interrogatives, etc.) that are housed in the CP domain of the clause. Complementizers have generally not constituted the focus of attention, but their existence and importance in establishing the limits of the left-peripheral layer cannot be denied. Indeed, the data from English and Spanish investigated in this paper show that complementizers like *that* and *que* are ubiquitous, and they can in fact co-occur with virtually all the left-peripheral constructions investigated in the literature so far. Of course, not all speakers from all dialects (of English or Spanish) accept the presence/absence of *that/que* with every CP-related phenomenon, and it is precisely that high degree of variation that the present paper has tried to draw attention to. Importantly, the foregoing discussion has revealed that, on occasion, it is not even clear that a non-high complementizer necessarily evinces the presence of a medial left-peripheral head (e.g., recomplementation as a TopicP phenomenon or as a biclausal phenomenon; see Section 3.2). If the seemingly medial (recomplementation) *that/que* and low ('jussive/optative') *that/que* turn out to be repeats of the high *that/que* in a bisentential/restart configuration (à la Villa-García and Ott 2022), then the total number of distinct complementizer realizations would actually be significantly reduced in both languages, despite superficial appearances.

Throughout our discussion, we have seen cases where the complementizer can by and large remain silent (e.g., high *that* in English, as discussed in Section 3.1), configurations in which both the complementizer and the left-peripheral phrase are realized concurrently (e.g., exclamatives with *vaya* plus *que* in certain varieties of Asturian Spanish, on which see Section 3.3), and patterns where only the left-peripheral constituent can be realized (namely wh-interrogatives in Peninsular Spanish; Section 3.4). In short, there are cases in which the presence of the complementizer is (i) optional, (ii) obligatory, or (iii) impossible. This state of affairs raises four major questions, presented in the introduction and repeated here. These questions now receive partial answers, pending further investigation of the constructions studied herein (and others):

(i) Is *that/que* a mere overt manifestation of an otherwise null/silent head (i.e., [$_{XP}$ YP [$_{X'}$ *that/que* vs. Ø . . .]])? If so, is *that/que* always present but deleted in the phonology (PF), ~~*that/que*~~, or else inserted when phonologically realized?

- The evidence adduced throughout suggests that merely reducing the presence/absence of *that/que* to a pronunciation parameter is rather simplistic. This would actually imply complete absence or obligatoriness in all contexts (assuming that the parameter is set to a particular value, not to either), contrary to fact, as amply demonstrated throughout. Put another way, the prospect of a unitary analysis that relies on PF realization seems untenable.

(ii) Is the presence of *that/que* indicative of more complex underlying structure instead?

- At least for cases including recomplementation, which have been convincingly analyzed recently as restarts in discourse camouflaging two underlying sentences, the answer to this question appears to be positive: multiple *that*s/*que*s are symptomatic of a more intricate syntactic configuration behind the scenes (see also Villa-García, in preparation, for the claim that exclamatives with *que* likewise involve a more elaborated structure behind the scenes than their *que*-less homologs).

(iii) Whatever the case may be, what determines the (non-)realization of the complementizer in different positions? Is it dialect-based? If so, are some dialects more prone to lexicalizing complementizers in different positions than others? Is complementizer lexicalization processing-based? Are there any other factors that play a role in determining the presence or absence of the complementizer, such as formal vs. informal contexts?

- In some cases, as we have seen, the complementizer is optional in some contexts, obligatory in certain environments, and mandatorily absent in others. This sometimes depends on the actual dialect in question. At this stage, it cannot be ascertained that a dialect omits complementizers more often than other dialects across left-peripheral constructions in either language. Regarding processing, the restart analysis of recomplementation (which, I argue, can be extended to subordinate 'jussive/optative' contexts) is compatible with this view, inasmuch as the restart contributes to facilitating the processing of the sentence. And lastly, factors such as formal vs. informal contexts do play a role in complementizer manifestation: for instance, omission of *that* in embedded declaratives in English has traditionally been attributed to colloquial registers; conversely, omitting *que* in requests is a feature characteristic of formal, written discourse in Spanish.

(iv) Does the presence vs. absence of *that/que* have a bearing on the semantics (LF)? Put another way, is complementizer realization just a PF matter, or are there LF reflexes as well?

- If we put aside processing, emphasis, and the formal vs. informal distinction, the cases explored herein do not seem to manifest meaning differences (regarding, e.g., the truth conditions of the sentences in question) depending on the presence/absence of *that/que*.

Further investigations into the constructions at issue will hopefully provide fuller answers to the major questions posed in this paper.

Be that as it may, the paper has tried to parameterize (and micro-parameterize) the different left-peripheral head realizations observed across English and Spanish. The major proposals in the literature to account for microvariation rely either on superficial PF realization or on features, and in addition to extant accounts, I have adumbrated an account that capitalizes on whether the specifier positions of the relevant maximal projections are filled overtly.

At present, however, a unified account of the facts across the board appears to be far from reachable. Many questions remain at this point regarding both the diatopic extent of the variation observed in relation to head realizations (especially in English, but also in Spanish) alongside the analysis of such variation in English and Spanish. It is my sincere hope that the data, dialectal contrasts, and lines of analysis gathered in this paper will contribute to this important enterprise.

Funding: This research was funded by by a María Zambrano International Talent Attraction Grant (MU-21-UP2021-030 71880965) at the LINGUO group of the University of Oviedo, awarded by the Spanish Ministry of Universities, with funding from the European Union (#NextGenerationEU, NGEU). Similarly, this project has benefited from the Spanish-Government-funded project INFOSTARS (PGC2018-093774-B-I00).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: I would like to thank three anonymous reviewers and the special editor, Ana Ojea, for their useful and thorough comments. I am also grateful to the English and Spanish consultants who helped me gather the relevant acceptability judgments. I would also like to express my gratitude to the conference abstract reviewers and/or the audiences at Ciudad Real (Castilla-La Mancha), Manchester, Bucharest, Geneva, Leiden, and Wake Forest (North Carolina). More specifically, I would like to thank the following individuals for their valuable observations: David Adger, Luigi Andriani, Delia Bentley, Ricardo Bermúdez-Otero, Giuliano Bocci, Željko Bošković, Ignacio Bosque, Simone De Cia, Antonio Cañas, Carlos Echeverría, Francisco Fernández-Rubiera, Daniel Á. García Velasco, Diego Gibanal Faro, Raquel González Rodríguez, Edita Gutiérrez, Vera Hohaus, Ángel Jiménez-Fernández, Bárbara Marqueta, Alexandru Nicolae, Isabel Pérez-Jiménez, Andrew Radford, Luigi Rizzi, Andrés Saab, Giuseppe Samo, Michelle Sheehan, Imanol Suárez-Palma, and Sam Wolfe.

Conflicts of Interest: The author declares no conflict of interest.

Notes

1 See also Uriagereka's (1995) FP system wherein FP is above CP.

2 Conversely, Llinàs-Grau and Bel (2019) reflect on corpus-based investigations whose results indicate that verbs such as *tell* and *hope* typically take a null-complementizer clausal complement, as opposed to *think*, *say*, and *know.*

3 The range of structures allowing this pattern, which is attested in formal varieties across dialects, is an open question that future research should care to address. For instance, the particular example in (22) appears to be a parenthetical use of *dicen*, as pointed out by an anonymous reviewer. As noted by a different anonymous reviewer, a revealing contrast in this regard would involve a sentence with subordination (e.g., a relative clause) and one without it, in which case omitting *que* would not be an option:

a.	*El*	*alumno*	*que*	*sabía*	*(que)*	*era*	*brillante.*
	the	student	that	knew	that	was	bright
	'The student who knew he was bright.'						
b.	**El*	*alumno*	*sabía*	*era*	*brillante.*		
	the	student	knew	was	bright		
	'The student knew he was bright.'						

4 An empirical problem with an analysis of *that*-less clauses as IPs/TPs is that there are data indicating that left-edge-related material can occur even in the absence of complementizers, as the following example demonstrates:

(i) *'I think ∅ the general physics community, ∅ they're a little bored with the equation,' he said* (*New York Times*, cited in Radford (2018, p. 111)).

5 For the sake of completeness, Spanish possesses a quotative marker (Etxepare 2010) in the shape of *que*:

(i)	*Que*	*se*	*han*	*convocado*	*elecciones*	*anticipadas.*
	that	cl.	have	summoned	elections	anticipated
	'Somebody/I said that a snap election has been called.'					

This element behaves syntactically like a high *que*, for it precedes wh-phrases, dislocations, and doubled complementizers (see Section 3.2) (Villa-García 2015). English lacks a quotative marker of this type, but does manifest *that* at the beginning of an utterance in clausal fragments, which for some speakers can be omitted in (ii)B:

(ii) A: *What did he say?*
B: *That you shouldn't count on him.*

6 In work in progress, I show that this is rather relevant to the analysis of inversion in Romance languages like Spanish. What the data indicate is that it is the full phrase containing the wh-exclamative and *que* that triggers inversion, not just the exclamative specifier. This leads us to conclude that even in the absence of the physical complementizer, the exclamative head is occupied by a null counterpart of *que*, since inversion occurs regardless of the presence of *que* (though see Villa-García, in preparation, for a dissenting view). Under some analyses (e.g., T°-to-C°), inversion is accounted for by assuming that the verb moves all the way to the head whose specifier hosts the exclamative phrase. However, the *que* data render this type of account implausible, since the head of the projection hosting the exclamative is occupied by the complementizer (unless additional projections are postulated).

7 In much the same way as in the case of recomplementation, different accounts of exclamatives have been proposed in the literature: CP (Bosque 1984; Brucart 1993, among others); FocusP (Hernanz and Rigau 2006); FocusP for the exclamative phrase

and Finiteness^{o} for *que* (Demonte and Fernández-Soriano 2009); ExclamativeP (Rizzi and Bocci 2017); the wh-exclamative is in a high CP and *que* in a lower CP (Zanuttini and Portner 2003; Tirado 2016; see also Benincà 1996); *qué* in ForceP, with the exclamative element in FocusP, and the low *que* in a lower-than-FocusP Topic^{o} (Gutiérrez-Rexach 2001); the *qué*-exclamative phrase moves from FocusP to ForceP, whose head is occupied by the complementizer, and the verb moves to Focus^{o}, accounting for inversion (Villalba 2019). Note, however, that inversion does not necessarily imply verb movement to the CP domain (viz. C^{o} as a null affix that needs adjacency with the verb, in the spirit of Buesa-García (2008)). In this sense, an anonymous reviewer poses the question of where the verb sits under the analysis in (38), since it is usually assumed that it moves all the way to the phrase whose specifier contains the exclamative constituent (see Villa-García, in preparation, for much relevant discussion).

Additionally, RAE-ASALE (2009) mentions the ongoing debate over whether such instances of *que* resemble relatives, an open question at present. The optionality of these complementizers seems to contravene the claim that they may be relatives, as *que* in Spanish relatives is mandatorily overt. RAE-ASALE notes that such exclamatives lack pied-piping, an expectation of the relative-clause analysis. Yet, sentences like the following, with the pied-pied preposition, are acceptable, suggesting that the issue is far from settled:

(i) ¡ *Vaya líos en los que te metes!*
vaya problems in the that cl. enter
'You (always) get involved in doggy stuff!'

8 Masullo (2017, p. 113) gives an example from Argentinean Spanish under *mirar* 'look' which contains *que* in brackets, pointing to the acceptability of this element in embedded contexts also in varieties other than in those varieties of Peninsular Spanish where pleonastic *que* is licit with *qué*-exclamatives:

(i) ¡ *Mirá qué bello (que) es el Nahuel Huapi!*
look what beautiful that is the Nahuel Huapi
'Look at how beautiful Nahuel Huapi is!'

9 However, two properties that further tease apart interrogatives and exclamatives is that interrogatives, but not exclamatives, can appear in long-distance configurations and can likewise occur in-situ.

10 A potential issue for this type of apparent counterexample is that *cuando termine* could be analyzed as an adjunct, which Villa-García (2015) shows can occur below FinitenessP.

11 Note in this connection the suggestion mentioned in Section 3.1 that *that*-less clauses may actually lack a CP altogether, which amounts to saying that clauses without *that* are not CPs (they are IPs/TPs) and therefore null *that* does not exist: the absence of *that* simply heralds the lack of the CP domain.

References

Adger, David. 2003. *Core Syntax: A Minimalist Approach*. Oxford: Oxford University Press.

An, Duk-Ho. 2007. Syntax at the PF Interface: Prosodic Mapping, Linear Order, and Deletion. Ph.D. dissertation, University of Connecticut, Storrs, CT, USA.

Antonelli, André. 2013. *Selected Proceedings of the 16th Hispanic Linguistics Symposium*. Edited by Jennifer Cabrelli Amaro, Gillian Lord, Ana de Prada Pérez and Jessi Elana Aaron. Somerville: Cascadilla Proceedings Project, pp. 15–26.

Bayer, Josef. 2014. Syntactic and phonological properties of wh-operators and wh-movement in Bavarian. In *Bavarian Syntax: Contributions to the Theory of Syntax*. Edited by Günther Grewendorf and Helmut Weiß. Amsterdam and Philadelphia: John Benjamins, pp. 23–50.

Bayer, Josef, and Probal Dasgupta. 2016. Emphatic topicalization and the structure of the left periphery: Evidence from German and Bangla. *Syntax* 19: 1–45. [CrossRef]

Benincà, Paola. 1996. La struttura della frase esclamativa alla luce del dialetto padovano. In *Italiano e Dialetti Nel Tempo: Saggi di Grammatica per Giulio Lepschy*. Edited by Paola Benincà, Guglielmo Cinque, Tullio De Mauro and Nigel Vincent. Rome: Bulzoni, pp. 23–43.

Bošković, Željko, and Howard Lasnik. 2003. On the distribution of null complementizers. *Linguistic Inquiry* 34: 527–46. [CrossRef]

Bošković, Željko. 1997. *The Syntax of Non-Finite Complementation: An Economy Approach*. Cambridge: MIT Press.

Bošković, Željko. 2016. On the timing of labeling: Deducing Comp-trace effects, the Subject Condition, the Adjunct Condition and tucking in from labeling. *The Linguistic Review* 33: 17–66. [CrossRef]

Bosque, Ignacio, ed. 2017. *Advances in the Analysis of Spanish Exclamatives*. Ohio: The Ohio State University.

Bosque, Ignacio. 1984. Sobre la sintaxis de las oraciones exclamativas. *Hispanic Linguistics* 1: 283–304.

Brovetto, Claudia. 2002. Spanish clauses without complementizer. In *Current Issues in Romance Languages: Selected Papers from the 29th Linguistic Symposium on Romance Languages*. Edited by Teresa Satterfield, Christina Tortora and Diana Cresti. Ann Arbor: LSRL, pp. 33–46. [CrossRef]

Brucart, José María. 1993. Sobre la estructura de SCOMP en español. In *Sintaxi: Teoria i Perspectives*. Edited by Amadeu Viana. Lleida: Amadeu Pagès, pp. 59–102.

Buesa-García, Carlos. 2008. *The Subject-Gap Restriction in Spanish Wh-Questions: Some Empirical and Theoretical Consequences*. Storrs: Second General Examination, University of Connecticut.

Casas, Antonio. 2004. Exclamativas y expletividad. El *que* enfático. *Revista de Filología Española* 84: 265–84. [CrossRef]

Cerrudo-Aguilar, Alba. 2014. *Contextos de (des)aparición de* que. Master's thesis, Autonomous University of Barcelona, Bellaterra, Catalonia, Spain.
Chomsky, Noam. 1981. *Lectures on Government and Binding. The Pisa Lectures*. Berlin and New York: De Gruyter Mouton. [CrossRef]
Chomsky, Noam. 1986. *Barriers*. Linguistic Inquiry Monograph Thirteen. Cambridge: MIT Press.
Chomsky, Noam. 1995. *The Minimalist Program*. Cambridge: MIT Press.
Chomsky, Noam, and Howard Lasnik. 1977. On Filters and Control. *Linguistic Inquiry* 8: 425–504.
Demonte, Violeta, and Olga Fernández-Soriano. 2009. Force and finiteness in the Spanish complementizer system. *Probus* 21: 23–49. [CrossRef]
Echeverría, Carlos. 2020. Recomplementation between Discourse and Grammar: A Syntactic Investigation Combining Experimental and Corpus Methods. Ph.D. dissertation, The Pennsylvania State University, State College, PA, USA.
Echeverría, Carlos, and Javier López Seoane. 2019. Recomplementation as constrained CP resetting. Paper presented at the Association of Linguistics and Philology of Latin America (Alfalito), Queens College, City University of New York, New York, NY, USA, September 20.
Escribano, José Luis Rafael G. 1991. *Una Teoría de la Oración*. Oviedo: Servicio de Publicaciones de la Universidad de Oviedo.
Espinal, M. Teresa, Cristina Real-Puigdollers, and Xavier Villalba. 2022. From a movement verb to an epistemic discourse marker. The diachronic change of Spanish *vaya*. *Language Variation*, 1–36. [CrossRef]
Etxepare, Ricardo. 1996. On Null Complementizers in Spanish. *Anuario Del Seminario De Filología Vasca "Julio De Urquijo"* 30: 469–96. [CrossRef]
Etxepare, Ricardo. 2010. From hearsay evidentiality to samesaying relations. *Lingua* 120: 604–27. [CrossRef]
Fontana, Josep M. 1993. Phrase structure and the syntax of clitics in the history of Spanish. Ph.D. dissertation, University of Pennsylvania, Philadelphia, PA, USA.
Frank, Joshua. 2020. An Experimental Approach to Recomplementation: Evidence from Monolingual and Bilingual Spanish. Ph.D. dissertation, University of Texas at Austin, Austin, TX, USA.
Franks, Steven. 2005. What Is That? In *Indiana University Working Papers in Linguistics*. Bloomington: Indiana University, vol. 5, pp. 33–62.
Green, Lisa. 2007. Syntactic Variation. *Sociolinguistic Variation: Theories, Methods, and Applications* 184: 24–44. Available online: https://scholarworks.umass.edu/linguist_faculty_pubs/184 (accessed on 20 August 2023).
Gutiérrez-Rexach, Javier. 2001. Spanish Exclamatives and the Semantics of the Left Periphery. In *Selected Papers from Going Romance*. Edited by Johan Rooryck, Yves de Hulst and Jan Schroten. Amsterdam and Philadelphia: John Benjamins, vol. 99, pp. 167–94.
Hegarty, Michael V. 1992. Adjunct Extraction and Chain Configurations. Ph.D. dissertation, MIT, Cambridge, MA, USA.
Henry, Alison. 1995. *Belfast English and Standard English: Dialect Variation and Parameter Setting*. Oxford: Oxford University Press.
Hernanz, Maria Lluïsa, and Gemma Rigau. 2006. Variación dialectal y periferia izquierda. In *Andolin Gogoan. Essays in Honour of Professor Eguzkitza*. Edited by Beatriz Fernández and Itziar Laka. Vitoria: Universidad del País Vasco, pp. 435–52.
Iatridou, Sabine, and Anthony Kroch. 1992. The licensing of CP-recursion and its relevance to the Germanic verb-second phenomenon. *Working Papers in Scandinavian Linguistics* 50: 1–25.
Koopman, Hilda. 1997. *The Doubly Filled C Filter, the Principle of Projection Activation and Historical Change*. Oakland: University of California.
Ledgeway, Adam. 2005. Moving through the left periphery: The dual complementizer system in the dialects of Southern Italy. *Philological Society* 103: 229–396. [CrossRef]
Ledgeway, Adam. 2020. The North-South Divide: Parameters of Variation in the Clausal Domain. *L'Italia Dialettale* 81: 29–77.
Llinàs-Grau, Mireia, and Aurora Bel. 2019. On the Acceptance of the Null That by Bilingual Catalan/Spanish L2 Learners of English. *Languages* 4: 30. [CrossRef]
López, Luis. 2009. Ranking the Linear Correspondence Axiom. *Linguistic Inquiry* 40: 239–76. [CrossRef]
Martínez-Vera, Gabriel. 2019. Phases, labeling, antilocality and intonational phrases: Recomplementation in Spanish. *Probus* 3: 187–231. [CrossRef]
Masullo, Pascual J. 2017. Exclamatives in (Argentinian) Spanish and Their Next of Kin. In *Advances in the Analysis of Spanish Exclamatives*. Edited by Ignacio Bosque. Ohio: The Ohio State University Press, pp. 139–58.
Oushiro, Livia. 2011. Wh-interrogatives in Brazilian Portuguese: The Influence of Common Ground. *University of Pennsylvania Working Papers in Linguistics* 17: 145–54.
Pesetsky, David, and Esther Torrego. 2001. T-to-C movement: Causes and consequences. In *Ken Hale: A life in language*. Edited by Michael Kenstowicz. Cambridge: MIT Press, pp. 355–426.
Plann, Susan. 1982. Indirect questions in Spanish. *Linguistic Inquiry* 13: 297–312.
Radford, Andrew. 2016. *Analysing English Sentences*, 2nd ed. Cambridge: Cambridge University Press. [CrossRef]
Radford, Andrew. 2018. *Colloquial English: Structure and Variation*. Cambridge: Cambridge University Press. [CrossRef]
RAE-ASALE. 2009. *Nueva Gramática de la Lengua Española*. Madrid: Espasa.
Rizzi, Luigi. 1997. The fine structure of the left periphery. In *Elements of grammar*. Edited by Liliane Haegeman. Dordrecht: Kluwer, pp. 281–337. [CrossRef]
Rizzi, Luigi. 2013. Notes on cartography and further explanation. *Probus* 250: 197–226. [CrossRef]

Rizzi, Luigi, and Giuliano Bocci. 2017. Left periphery of the clause: Primarily illustrated for Italian. In *The Wiley Blackwell Companion to Syntax*, 2nd ed. Edited by Martin Everaert and Henk van Riemsdijk. Hoboken: John Wiley & Sons, pp. 1–30. [CrossRef]
Roberts, Ian, and Anna Roussou. 2003. *Syntactic Change: A Minimalist Approach to Grammaticalization*. Cambridge: Cambridge University Press. [CrossRef]
Rodríguez-Ramalle, Teresa. 2003. *La gramática de los adverbios terminados en –mente o cómo expresar maneras, opiniones y actitudes a través de la lengua*. Madrid: Ediciones de la Universidad Autónoma de Madrid.
Rodríguez-Riccelli, Adrián. 2018. Espero estén todos: The distribution of the null subordinating complementizer in two varieties of Spanish. In *Language Variation and Contact-Induced Change: Spanish across Space and Time*. Edited by Jeremy King and Sandro Sessarego. Amsterdam and Philadelphia: John Benjamins, pp. 299–332. [CrossRef]
Silva-Corvalán, Carmen. 1994. *Language Contact and Change: Spanish in Los Angeles*. Oxford: Oxford University Press.
Smith, Jennifer, and David Adger. 2005. Variation and the minimalist program. In *Syntax and Variation: Reconciling the Biological and the Social*. Edited by Leoni Cornips and Karen P. Corrigan. Amsterdam and Philadelphia: John Benjamins, pp. 149–78. [CrossRef]
Subirats-Rüggeberg, Carlos. 1987. *Sentential Complementation in Spanish: A Lexico-Grammatical Study of Three Classes of Verbs*. Amsterdam and Philadelphia: John Benjamins.
Thoms, Gary, David Adger, Caroline Heycock, and Jennifer Smith. 2019. Syntactic variation and auxiliary contraction: The surprising case of Scots. *Language* 95: 421–55. [CrossRef]
Tirado, Irene. 2016. Aproximación sintáctica a las oraciones exclamativas del español. *Borealis—An International Journal of Hispanic Linguistics* 5: 191–219. [CrossRef]
Torrego, Esther. 1983. More effects of successive cyclic movement. *Linguistic Inquiry* 14: 561–65.
Uriagereka, Juan. 1995. An F position in Western Romance. In *Discourse configurational Languages*. Edited by Katalin Kiss. Oxford Studies in Comparative Syntax. Oxford: Oxford University Press, pp. 153–75.
Villa-García, Julio. 2012. Recomplementation and locality of movement in Spanish. *Probus* 24: 257–314. [CrossRef]
Villa-García, Julio. 2015. *The Syntax of Multiple-Que Sentences in Spanish: Along the left periphery*. Issues in Hispanic and Lusophone Linguistics. Amsterdam and Philadelphia: John Benjamins, vol. 2. [CrossRef]
Villa-García, Julio. 2018. Syntactic microcomparison in the Complementizer Phrase domain: C° realisations across Spanish. Paper presented at the 2nd Spanish Dialects Meeting (II Spadisyn), University of Castille-La Mancha, Ciudad Real, Spain, May 17–18.
Villa-García, Julio. 2019. Recomplementation in English and Spanish: Delineating the CP Space. *Glossa: A Journal of General Linguistics* 4: 1–44. [CrossRef]
Villa-García, Julio. 2023. Hanging Topic Left Dislocations as extrasentential constituents: Toward a paratactic account. Evidence from English and Spanish. *The Linguistic Review* 40: 265–310. [CrossRef]
Villa-García, Julio. In preparation. Lessons from Exclamatives in Spanish Varieties: Implications for the Account of Obligatory Subject-Verb Inversion. Oviedo: University of Oviedo. Manchester: University of Manchester, *in preparation*.
Villa-García, Julio, and Dennis Ott. 2022. Recomplementation as a paratactic phenomenon: Evidence from Spanish and English. *Journal of Linguistics* 60: 1–32. [CrossRef]
Villa-García, Julio, and Raquel González Rodríguez. 2020. Dialectal variation in the expression of *que* in *sí-que* 'yes that' contexts across Spanish: The case of some Latin American Spanish varieties. *Glossa: A journal of General Linguistics* 5: 1–20. [CrossRef]
Villalba, Xavier. 2019. Romance wh-exclamatives: A view from the periphery. Paper presented at the 29th Colloquium on Generative Grammar, University of Castille-La Mancha, Ciudad Real, Spain, May 23–25.
Vincent, Nigel. 2019. CP and COMP in Diachrony. In *Proceedings of the LFG'19 Conference, Australian National University*. Edited by Miriam Butt, Tracy Holloway King and Ida Toivonen. Stanford: CSLI Publications, pp. 314–33.
Watanabe, Akira. 1996. *Case Absorption and Wh-Agreement*. Dordrecht: Kluwer.
Zanuttini, Raffaella, and Paul Portner. 2003. Exclamative Clauses: At the Syntax-Semantics Interface. *Language* 79: 39–81. [CrossRef]
Zwicky, Arnold. 2002. I Wonder What Kind of Construction That This Example Illustrates. In *The Construction of Meaning*. Edited by David I. Beaver, Luis D. Casillas Martinez, Brady Z. Clark and Stefan Kaufmann. Chicago: The University of Chicago Press.

MDPI AG
Grosspeteranlage 5
4052 Basel
Switzerland
Tel.: +41 61 683 77 34

Languages Editorial Office
E-mail: languages@mdpi.com
www.mdpi.com/journal/languages

Academic Open
Access Publishing

mdpi.com

ISBN 978-3-7258-3593-5