

# BEYOND PHILOLOGY

AN INTERNATIONAL JOURNAL  
OF LINGUISTICS, LITERARY STUDIES  
AND ENGLISH LANGUAGE TEACHING

18/3

CONTRASTIVE LINGUISTICS

Edited by Javier Fernández-Sánchez

WYDAWNICTWO UNIWERSYTETU GDAŃSKIEGO  
GDAŃSK 2021

**EDITOR-IN-CHIEF**

Danuta Stanulewicz

**SECTION EDITORS**

Olga Sokołowska (Linguistics)

Ludmiła Gruszewska-Blaim (Literary Studies, Culture)

Olga Kubińska (Translation)

Anna Bączkowska (Language Acquisition, Academic Teaching)

Magdalena Wawrzyniak-Słiwska (Education)

Jadwiga Węgrodzka (Reviews, Reports, Interviews)

**ASSISTANT EDITORS**

Maria Fengler

Marcin Fortuna

Michał Golubiewski

Ewelina Gutowska-Kozielska

Karolina Janczukowicz

Karol Plichta

Joanna Redzimska

Małgorzata Smentek

**EDITORIAL ASSISTANT**

Aleksandra Słyszewska

**PROOFREADERS**

Martin Blaszk

Sarah Flamminio

Jean Ward

Tadeusz Z. Wolański

The Editors and Proofreaders are all affiliated with the Institute of English and American Studies, University of Gdańsk, Poland.

**COVER DESIGN**

Andrzej Taranek

**ISSN 1732-1220**

**eISSN 2451-1498**

© Copyright by Uniwersytet Gdański

Wydawnictwo Uniwersytetu Gdańskiego

**Contact address**

Institute of English and American Studies

University of Gdańsk

ul. Wita Stwosza 51

80-308 Gdańsk

Poland

Phone: (+48) 58 523 30 49, (+48) 58 523 30 50

Email: [beyond.philology@ug.edu.pl](mailto:beyond.philology@ug.edu.pl)

## **ASSOCIATE EDITORIAL BOARD**

Marta Bogdanowicz (University of Gdańsk, Poland)  
Joanna Burzyńska-Sylwestrzak (Uczelnia Lingwistyczno-Techniczna,  
Świecie, Poland)  
Ewa Dąbrowska (Northumbria University, Newcastle, U.K.)  
Irene Gilsenan Nordin (Dalarna University, Falun, Sweden)  
Desmond Graham (University of Newcastle, U.K.)  
Gabriela Kelemen (Aurel Vlaicu University of Arad, Romania)  
Zoltán Kövecses (Eötvös Loránd University, Budapest, Hungary)  
Ronald W. Langacker (University of California at San Diego, U.S.A.)  
Barbara Lewandowska-Tomaszczyk (University of Łódź, Poland)  
David Malcolm (SWPS University, Warszawa, Poland)  
Elżbieta H. Oleksy (University of Łódź, Poland)  
Adam Pasicki (Pedagogical University of Kraków, Poland)  
Piotr Ruszkiewicz (Pedagogical University of Kraków, Poland)  
Otilia Costa e Sousa (Instituto Politécnico de Lisboa, Portugal)  
Bogdan Szymanek (Catholic University of Lublin, Poland)  
Ryszard Wenzel (Akademia Polonijna, Częstochowa, Poland)  
Marta Wiszniewska (Nicolaus Copernicus University, Toruń, Poland)

## **BOARD OF REVIEWERS**

Frank Cioffi (Baruch College, City University of New York, U.S.A.)  
Anna Cisło (University of Wrocław, Poland)  
Artur Czapiga (University of Rzeszów, Poland)  
Tadeusz Danilewicz (Wyższa Szkoła Bankowa, Gdańsk, Poland)  
Danuta Gabryś-Barker (University of Silesia, Katowice, Poland)  
Aleksandra Kędzierska (Maria Curie-Skłodowska University, Lublin,  
Poland)  
Ewa Komorowska (University of Szczecin, Poland)  
Wojciech Kubiński (Państwowa Wyższa Szkoła Zawodowa w Elblągu, Poland)  
Elżbieta Mańczak-Wohlfeld (Jagiellonian University, Kraków, Poland)  
Marzenna Mioduszevska (Independent Scholar, Madrid, Spain)  
Grzegorz Moroz (University of Białystok, Poland)  
Wojciech Nowicki (Uczelnia Państwowa im. Szymona Szymonowica  
w Zamościu, Poland)  
Kazimierz Sroka (University of Gdańsk, Poland; Polonia University in  
Częstochowa, Poland)  
Krystyna Stamirowska (Jagiellonian University, Kraków, Poland)  
Yuri Stulov (Minsk State Linguistic University, Belarus)  
Kamila Turewicz (Państwowa Wyższa Szkoła Zawodowa w Płocku, Poland)  
Agnieszka Uberman (University of Rzeszów, Poland)  
Jacek Waliński (University of Łódź, Poland)  
Tomasz Warchoń (Georgia Southern University, Statesboro, U.S.A.)  
Julian Wolfreys (Independent Scholar, U.K.)

## **ISSUE REVIEWERS**

Isabel Crespí (Queen Mary University London, U.K.)  
Hortènsia Curell Gotor (Universitat Autònoma de Barcelona, Spain)  
David Erschler (Ben Gurion University of the Negev, Israel)  
Alfredo García-Pardo (State University of New York,  
Francisco Javier Martín González (Universidad de Sevilla, Spain)  
Lucia Medea-García (University of Lisbon, Portugal)  
Mireia Llinàs-Grau (Universitat Autònoma de Barcelona, Spain)  
Ares Llop Naya (University of Oxford, U.K.)  
Paolo Lorusso (Università degli Studi di Udine, Italy)  
Anna Paradís (University of Cambridge, U.K.)  
Susagna Tubau (Universitat Autònoma de Barcelona, Spain)

*Beyond Philology* is published in print and online:

<[https://fil.ug.edu.pl/wydzial\\_filologiczny/instituty/institut\\_anglistyki\\_i\\_amerykanistyki/czasopismo\\_naukowe\\_beyond\\_philology](https://fil.ug.edu.pl/wydzial_filologiczny/instituty/institut_anglistyki_i_amerykanistyki/czasopismo_naukowe_beyond_philology)>,

<<https://czasopisma.bg.ug.edu.pl/index.php/beyond>>.

Licence: CC-BY

The online version is primary.

*Beyond Philology* is indexed by

- The Central European Journal of Social Sciences and Humanities
- ERIH PLUS European Reference Index for the Humanities and Social Sciences
- Index Copernicus
- MLA International Bibliography
- Norwegian Register for Scientific Journals, Series and Publishers

## **BEYOND PHILOLOGY 18/3**

### **Contents**

The internal structure of proper names: Surnames, patronymics and relational elements ANTONIO FÁBREGAS	7
Non-verbal complements of modal verbs: The case of directional adverbs in Czech HANA GRUET-SKRABALOVA	45
On wh-movement in echo questions and crosslinguistic variation EKATERINA CHERNOVA	65
On the presence and absence of the complementizer in embedded gapping in Spanish, Polish and English JAVIER FERNÁNDEZ-SÁNCHEZ	109
How language shapes interpersonal distance: Analysis of pronominal forms of address in Spanish, Polish and Italian MAREK BARAN EWA URBANIAK	137
Uniqueness in languages with and without articles: Catalan vs Russian DARIA SERÉS JOAN BORRÀS-COMES M. TERESA ESPINAL	163

Definiteness in second language acquisition: Preliminary results regarding Indo-European and Afroasiatic languages CARLA FERRERÓS PAGÈS FRANCESC ROCA URGELL MARTA VILOSA SÁNCHEZ	197
Information for Contributors	239

**The internal structure of proper names:  
Surnames, patronymics  
and relational elements**

ANTONIO FÁBREGAS

*Received 15.02.2021,  
accepted 17.10.2021.*

**Abstract**

This article researches patronymics in a broad sense – taken as components of a proper name that, morphologically, can be decomposed in a first name and a morpheme – with a focus on Spanish and Belarusian – the second conforming to a narrow definition of patronymic, where it is a component of a proper name distinct from both the first name and the surname. Our claim is that patronymics are the syntactic result of combining a first name with relational structure, a PP layer in the case of Spanish, and both a pP and a PP layer in the case of Belarusian. This research will allow us to probe inside the internal structure of complex proper names, including the relation between first name and surname, first name and patronymic, complex first names and complex surnames.

**Keywords**

proper names, patronymics, prepositions, reference, comparative linguistics

## **Wewnętrzna struktura imion własnych: nazwiska, patronimika i elementy relacyjne**

### **Abstrakt**

Ten artykuł dotyczy patronimiki w szerokim znaczeniu, rozumianej jako elementy składowe imienia własnego, które morfologicznie można rozłożyć na imię i morfem, ze szczególnym uwzględnieniem języka hiszpańskiego i białoruskiego. W wąskim znaczeniu patronimika jest składnikiem imienia własnego, odrębnego zarówno od imienia, jak i nazwiska. Głównym twierdzeniem jest to, że patronimika jest syntaktycznym wynikiem połączenia imienia ze strukturą relacyjną; frazą przyimkową w przypadku języka hiszpańskiego, a w przypadku białoruskiego pP (small p phrase) oraz PP. Badania te pozwolą nam zagłębić się w wewnętrzną strukturę złożonych imion własnych, w tym w relacje między imieniem a nazwiskiem, imieniem a patronimem, złożonymi imionami i złożonymi nazwiskami.

### **Słowa kluczowe**

nazwy własne, patronimiki, przyimki, odniesienie, językoznawstwo kontrastywne

### **1. Introduction**

Despite the attention paid to proper names in formal semantics (Kripke 1980, Evans 1982, Kaplan 1989, Recanati 1997) and syntax (Longobardi 1994; see also Abbott 2002, Matushansky 2008, Sainsbury 2015), surprisingly little has been said about the internal structure of proper names, and more specifically about the type of heads and configurations that combinations of proper names within the same constituent contain. In a language like Spanish, English or Norwegian, human proper names can be simple (1) or complex (2), and when they are complex they typically correspond to what is socially called 'first name plus surname' (2a), combinations of



two first names into a complex one (2b) or combinations of two surnames into a complex surname (2c).

(1) *Pedro, Salvador, María, Federico, Marta, Luis...*

(2) a. *Salvador Dalí*

b. *José María, María José, Pedro Pablo...*

c. *García Lorca*

The question that immediately arises is whether all combinations in (2) are of equal status, and what type of configurations they represent.

In this article we will pay particular attention to cases of complex proper names where the second element carries a designated suffix that marks it necessarily as a surname or another type of complement or modifier of the first name. For the case of Spanish this involves the suffix *-ez*, which was used to create so-called surnames from the first name of the father (3). A relevant point of comparison is the so-called patronymic in Slavic, here and in the rest of the article illustrated with Belarusian, which is also characterized by a specific morpheme (*-avich / -yevich* for the masculine) which attaches to the first name of the father (4). There are several dimensions where the Spanish suffix and the Belarusian one differ, and we will review these in due course, but the point of interest here is that in both cases we have a decomposable morpheme that is used to introduce the second member of a complex proper name.

(3) *Fernández* → *Fernánd-ez*

(4) *Aliaksandr* → *Aliaksandr-auna, Aliaksandr-avich*

The main question is the following: what is the nature of these suffixes, what type of head and configuration underlies to them, and as a result what type of complex proper name they produce? We will argue, in fact, that these suffixes correspond to relational structures that at a certain level of abstraction are identical: truncated prepositional phrases which produce a complex constituent that is later on nominalized through little *n*.

The structure of this paper is as follows. In the next section (§2) we will provide a description of patronymic and patronymic-like members within complex proper names, where we will also discuss the main asymmetries between the Spanish and the Belarusian affixes. In §3 we will provide a more detailed description of the syntactic behavior of the first name + surname combinations in Spanish, and in §4 we will do the same with the Belarusian patronymic. Section §5 provides the analysis, which is divided in two parts: the aspects that are shared by the two languages and the specificities that differentiate them, which ultimately reduce to the presence or absence of uninterpretable phi features in the case of Belarusian.

## 2. Morphological patronymics in Spanish and otherwise

Patronymics can be defined both in a broad sense and in a narrower sense, of which the broader sense will be the one that is relevant in this article. In the broad sense, a patronymic is a human proper name derived from the first name of one of the parents, typically the father (Kohlheim and Kohlheim 2000, van Langendonck 2007). Crucially, the patronymic in this broad sense is always a component of a complex proper name which cannot be used as a first name. The patronymic in this broad sense has the morphological shape in (5), depending on whether the morpheme acts as a suffix or as a prefix.

- (5) a. [[first name] morpheme]  
 b. [morpheme [first name]]

This broad sense, from now on called 'morphological patronymic', allows the structure in (5) to be the so-called surname, but it does not force it to be. In the narrow sense, the patronymic is one of the three components of the basic human proper name in many Slavic languages, in contrast both to the first name and the surname. The social conditions of usage of

patronymics in this narrow sense are different from those involving first names or surnames; using the first name and the patronymic without the surname is normally associated to a high degree of respect –e.g., students to teachers–, and it is common that the morphemes used for narrow patronymics are different from those used to derive surnames (e.g., Russian *Ivanovich* vs. *Ivanov*).

A Slavic surname that is derived from a human first name, like *Ivan-ov*, is a patronymic in the broad sense used in this article, a morphological patronymic, even if it is clearly distinct in morphological shape and social usage from the patronymic *Ivan-ovich* in the narrow sense. From now on, when we refer to patronymics the reader should interpret that with them we mean the broad sense of the term, those that correspond to the structures in (5).

Even though the goals of this article do not include attempting to present a typological overview of patronymics, a few examples are relevant in order to define two main dimensions of grammatical behavior that we will consider in our analysis. Cross-linguistically, patronymics are typically built through affixes (6) or kinship terms corresponding or etymologically related to the nouns that denote the descendants of someone, such as 'son' and 'daughter' (7).

(6) a.	<i>Adán</i>	→	<i>Adán-ez</i>	Spanish
	Adam		Adam-PTR	
b.	<i>Aliaksander</i>	→	<i>Aliaksandr-avich</i>	Belarusian
	Alexander		Alexander- PTR.MASC	
c.	<i>Owen</i>	→	<i>B-owen</i>	Welsh
	Owen		PTR-Owen	
d.	<i>Kowal</i>	→	<i>Kowal-ski</i>	Polish
	Kowal		Kowal-PTR	
e.	<i>Yousaf</i>	→	<i>Yousaf-zai</i>	Afghan
	Yousaf		Yousaf-PTR	
f.	<i>Donald</i>	→	<i>Mac-Donald</i>	Irish
	Donald		PTR-Donald	
(7) a.	<i>Peder</i>	→	<i>Peder-sen</i>	Norwegian
	Peder		Peder-son.of	

b. <i>Johan</i>	→	<i>Johan-s-son</i>	Swedish
Johan		Johan-of-son	
c. <i>Jón</i>	→	<i>Jón-s-son</i>	Icelandic
Jón		Jón-of-son	

There are other attested options, which can co-occur with the procedures described above in particular languages. A typical case is to build a patronymic through the use of a functional preposition roughly corresponding to 'of', as can be done for instance in Portuguese (*Da Silva* 'child of Silva'). In examples (7b) and (7c) above the internal morphological shape of the patronymic shows an additional linking element -s- which plausibly corresponds to this type of relational element.

There are two factors that differentiate the grammatical behavior of morphological patronymics across languages. The first one is sensitivity to the gender of the referent that carries that proper name. In a language like Norwegian or Swedish, there is no gender sensitivity, which means that the patronymic is built by adding -son / -sen independently of whether the person carrying that surname is male or female. In contrast, systematically, Slavic patronymics –now in the narrow sense, that is, as opposed to surnames– differentiate gender of the referent by masculine or feminine endings of the suffix.<sup>1</sup>

(8) <i>Aliaksandr-avich</i>	→	<i>Aliaksandr-auna</i>
Alexander- PTR.MASC		Alexander-PTR.FEM

Icelandic also makes this differentiation: from the same father called Jón, his male sibling would carry a surname built with -sson, while his daughter would carry a surname built with -dóttir, obviously related to the word for 'daughter'.

---

<sup>1</sup> An anonymous reviewer points out an important caveat to this generalization: in some languages where agreement takes place, like Polish, when the surname is adapted to another language, for legal reasons it becomes fixed in gender (e.g., *Monica Lewinski*, not *Monica Lewinska*).

(9) <i>Jón-s-son</i>	→	<i>Jón-s-dóttir</i>
Jón-of-son	→	Jón-of-daughter

The second broad distinction is the compositionality of meaning. In some languages, the patronymic is compositionally derived from the name of the male parent; in examples (8) and (9) above the person carrying that proper name must necessarily have a father called Aleksander or Jón: the meaning of the patronymic is compositionally decomposable through the morphemes. In contrast, in other languages, like Norwegian, Icelandic or Irish, being called Jonsson or MacDonald does not allow us to conclude that the father of that person is called Jon or Donald.

Even though we have not gathered a rich enough data set representing enough languages, a preliminary observation that can be made is that – at least for the languages considered here – the two properties partially correlate: languages whose patronymics are semantically decomposable (that is, where N-patronymic means 'child of N') are languages where the patronymic is sensitive to the gender of the referent. We have not found languages where the patronymic is assigned arbitrarily without reference to the first name of the actual father and the resulting proper name does not take gender into account, although not having a lot of languages we would not feel comfortable claiming that this is necessarily a typological generalization. It is, however, strong enough to be worth considering as part of the explananda in our analysis, as we will see in section §5.

It is crucial for our purposes to show that even if the patronymic is not semantically transparent it still has an internal morphological structure, that is, that it can be decomposed. Let us discuss this concentrating now (and from now on, in the rest of the article) in the case of Spanish.

Spanish patronymics are traditionally called 'surnames', and they are composed of a base that corresponds to a human male first name and the suffix *-ez*. A list of some of the most unusual surnames following this morphological pattern is pro-

vided in (10), where the original proper name is given in parenthesis.

- (10) a. *Fernandez (Fernando), Álvarez (Álvaro), Martínez (Martín), Sánchez (Sancho), Íñiguez (Íñigo), González (Gonzalo), Rodríguez (Rodrigo), Garcés (García), Benítez (Benito), Jiménez (Jimeno), Domínguez (Domingo), Antóniez (Antonio), Antolínez (Antolín), Adánez (Adán), Javiérez, Javier), Bernárdez (Bernardo), Diéguéz (Diego), Márquez (Marcos), Segúndez (Segundo)...*
- b. *Pérez (Pero ~ Pedro), Élez (Elio), Ibáñez (Iván), Peláez (Pelayo)*
- c. *Gómez (Gumo), López (Lope), Velásquez ~ Velázquez (Velasco ~ Velazco), Gutiérrez (Gutierre), Suárez (Suaro ~ Suero)*

The patronymics in (10a) have bases that are easily recognized as proper names in contemporary Spanish; those in (10b) contain either non unusual versions of existing proper names, or the addition of the suffix has produced some changes on the base that make reconstructing the proper name difficult. Those in (10c) are also derived as patronymics, but from bases that are no longer used generally as first names in contemporary Spanish.

My claim is, however, that in all cases speakers identify the word as a patronymic, independently of whether they identify the base as a first name or not, and even though the meaning of the patronymic is not compositional. The next few paragraphs present my arguments for this claim.

The surnames that have the patronymic morphological shape in Spanish – see below, section §3, for those surnames that lack this morphological shape – are not sensitive to the gender or the referent and non-semantically decomposable, but I will claim that they are still segmented as derived in the minds of contemporary speakers, even when the base proper noun is difficult to identify. My first argument for this claim is that, trivially, all surnames in (10) are systematically charac-

terized by the same, phonologically stable sequence, final *-ez*. Even when the base that results from decomposing this final sequence is not directly identifiable, speakers do recognize that *-ez* is a segmentable unit that is associated with a stable set of grammatical and semantic properties, specifically indicating 'formation of a surname'.

The second argument is related to this: contemporary speakers, in humorous styles, can produce neologisms that contain *-ez* from different types of bases, to create surnames that characterize someone by their properties. Among the examples that can be easily documented are those in (11).

(11) a.	<i>puta</i>	→	<i>Pút-ez</i>
	whore, bitch		bitch-PTR (cf. <i>hijo de puta</i> , 'son of a bitch')
b.	<i>cretino</i>	→	<i>Cretín-ez</i>
	idiot		idiot-PTR
c.	<i>bestiajo</i>	→	<i>Bestiáj-ez</i>
	beast		beast-PTR

These neologisms show that (i) speakers, even in contemporary Spanish, where the patronymics are not semantically decomposable, segment a morpheme *-ez* and (ii) that this morpheme is clearly associated to building surnames.

My third argument has to do with the phonological transparency of the patronymics in Spanish. Unlike most derivational suffixes, *-ez* is special in systematically keeping the stress of the base. When the base has the stress in the second-to-last syllable, which is the most frequent case, the patronymic keeps stress in that same syllable (12a). When the base is proparoxytone, the stress is also kept in that syllable also in the patronymic (12b). If the first name ends in a stressed syllable, the suffix is added to it without shifting stress (12c), and when the base has stress on a final or almost final vowel that is cancelled phonologically when *-ez* is added, stress falls on the last syllable, that is, on *-ez* itself (12d, where avoidance of

two interdental consonants leads to dissimilation of the final segment).

- (12) a. *M[á]rcos* → *M[á]rquez*, *Dom[í]ngo* → *Dom[í]nguez*  
 b. *[Á]lvaro* → *[Á]lvarez*, *[Í]ño* → *[Í]ñiguez*  
 c. *Ad[á]n* → *Ad[á]nez*, *Antol[í]n* → *Antol[í]ñez*  
 d. *Garc[í]a* → *Garc[é]s*

Tellingly, there is at least one case which, although historically related to *-ez*, does not keep *-ez* on the surface for phonological reasons: *Muñoz*, from the old proper name *Munio*, stressed (as far as we know) as *M[ú]nio*. This case, where *-ez* is not visible on the surface, is exceptional in carrying stress in a different syllable from the base noun, also lost in contemporary Spanish: *Muñ[ó]z*, not \**M[ú]ñoz*. There is then nothing phonological that can explain why those surnames derived with *-ez* keep stress on the same syllable as the base, rather the opposite: as it is generally the case in Spanish, a word ending in a consonant prefers to carry stress on the last syllable. Segmentation of *-ez* is what explains that stress is preserved on the base in contemporary Spanish.

My fourth and final argument in favor of decomposing the surnames in (10) into morphemes, despite the lack of semantic motivation, is the fact that nouns that contain *-ez* can never behave as first names. In Spanish, there is a lexical test that differentiates between first names and surnames, at least in formal registers: the honorific *don* / *doña* can only be attached to first names, while the title *señor* / *señora* can only be combined with surnames. Outside from very stigmatized sociolects, combinations like those in (13c,d) are out, in contrast with those in (13a,b); we mark with '%' the fact that the combinations in (13c,d) are only acceptable in those sociolectal varieties.

- (13) *Salvador Dalí*  
 a. *Don Salvador*  
 HONORIFIC *Salvador*



- b. *Señor Dalí*  
Mister Dalí
- c. %*Señor Salvador*
- d. %*Don Dalí*

Importantly, with proper names that contain *-ez* the test shows that they must be used as surnames and are blocked as proper names.

- (14) a. %*Don Martínez*  
b. *Señor Martínez*

This is not always the case with proper names without this morpheme. Social conventions definitely define some proper names as most frequently used as surnames or first names, but it is easy to find cases of proper names that can be used as both. In fact, *Martín*, the base for the patronymic in (15), is one such case. A male could be called *Martín Martín*, the first as birth name and the second as surname, but nobody can have as first name *Martínez*, which must be a surname.

- (15) a. *Don Martín*  
b. *Señor Martín*

Speakers, then, identify that human proper names with *-ez* are surnames, which is another argument to say that even when the base is not identifiable the role of the suffix is identified within the formation. The obvious question at this point is what happens with surnames that have no identifiable *-ez* ending, and specifically whether these are different or not from the ones where the patronymic is built overtly. Our claim will be that, as surnames, these behave also as the patronymics with *-ez*, and the reason is that in them the internal structure is identical but the head corresponding to the patronymic is not represented by a separate exponent. But before we argue for this, let us look deeper at the internal structure of proper names in Spanish.

### 3. The internal structure of human proper names in Spanish

As was already mentioned in the introduction, human proper names can be complex in three senses, which now we can be more precise about. First, a human proper name can correspond to the combination of proper names of different status: first name and surname, with relevant for instance for Belarusian, patronymic (16); a combination of two first names (17) or a combination of two surnames (18).

- |      |  |            |
|------|--|------------|
| (16) | a. <i>Felipe</i> <sup>first name</sup> <i>González</i> <sup>surname</sup>  | Spanish    |
|      | b. <i>Aliena</i> <sup>first name</sup> <i>Aliaksandrauna</i> <sup>patronymic</sup><br><i>Ramancuk</i> <sup>surname</sup> | Belarusian |
| (17) | a. <i>Camilo José</i><br>b. <i>José María</i><br>c. <i>María José</i><br>d. <i>Gloria Camila</i>                         |            |
| (18) | a. <i>Álvarez Martínez</i><br>b. <i>Ibáñez Serrador</i><br>c. <i>Fernández Sánchez</i>                                   |            |

The questions that we want to address for Spanish in this section are the following: (a) are the combinations in (17) and those in (18) of the same status or not?; (b) what type of relation is established between the members in (16a), which we have seen are differentiated at least by the combination of *don* / *señor*? This section will present the facts for Spanish, which will be the base of our analysis in §5.

Let us first determine the distribution of patronymic surnames in Spanish, starting with tests that show that they are properly classified as proper names – something that might be trivial in Spanish, but will not be so when we confront them with Belarusian patronymics. Crucially, surnames can be used in Spanish as proper names in the absence of a first name.

- (19) *Gómez*            *llegó*                            *anoche*.  
          *Gómez*            arrive.PAST.3SG            yesterday-night  
          ‘Gómez arrived yesterday evening.’

The context, of course, must be one where not mentioning the first name is enough to identify the referent, but (19) displays one of the crucial properties of proper names in Spanish: the possibility of appearing as preverbal subjects without any overt determiner or quantifier. The contrast in (20) shows that this is impossible with common nouns, but possible with first names, in sharp contrast to some Germanic languages where determiner-less common nouns can appear in preverbal subject position under certain conditions.

- (20) a.\* *Profesora*            *llegó*                            *anoche*.  
          professor            arrive.PAST.3SG            yesterday-night  
          Int.: ‘The professor arrived yesterday evening.’  
       b. *Marta*                    *llegó*                            *anoche*.  
          Marta            arrive.PAST.3SG            yesterday-night  
          ‘Marta arrived yesterday evening.’

Second, with proper names the combination with qualifying adjectives has two effects (Longobardi 1994): the adjective must appear prenominal, never postnominally (21, unless of course reinterpreted as part of the proper name, therefore losing its predicate status), and a determiner must be used in combination with the proper name (22) when used as an argument.

- (21) a. *Mi*                    *querido*                    *Pedro*.  
          my                    beloved                    Pedro  
       b.\**Mi*                    *Pedro*                    *querido*.  
          my                    Pedro                    beloved  
       (22) a.\**Querido Pedro* *llegó*                            *anoche*.  
          beloved Pedro arrive.PAST.3SG yesterday-night  
       b. {*El/ Mi*} *querido Pedro* *llegó anoche*.

The same pattern of data emerges with surnames, when used without proper names in the contexts mentioned for example in (19).

- (23) a. *Mi querido Álvarez*  
           my beloved Álvarez  
       b.\**Mi Álvarez querido*  
           my Álvarez beloved  
       c.\**Querido Álvarez llegó anoche.*  
           beloved Álvarez arrive.PAST.3SG yesterday-night

Thus, the tests tell us that (unsurprisingly) patronymics in Spanish are proper names, just like first names. However, this does not mean that they are the same type of proper names, as suggested by a set of asymmetries between them that we believe have not been described in the literature before.

The asymmetries – beyond the morphological shape, where *-ez* proper names must always be used as surnames – emerge particularly when one considers combinations of two first names or two surnames.

Both sequences of two first names and sequences of two surnames can be built in Spanish, as we have seen. However, despite appearances, the combination of two (or more) first names is not equivalent to the combination of two surnames, as a number of asymmetries show.

- (24) a. *José María*  
       b. *Fernández Álvarez*

Let us start with the prosodic properties of first names and surnames. Consider a sequence like (25), where the intermediate proper name *-Martín-* is one that can be used both as a first name and a surname.

- (25) *Luis Martín Álvarez*

Prosody differentiates between the structure where *Martín* is part of the first name and the one where it is part of the surname. In the first case, there is an intonational break after *Martín* and the first proper name is deaccented (26a). In the second case, the intonational break is after *Luis*, carrying stress, and there is a second intonational break after *Martín*. None of the proper names gets deaccented.

- (26) a. (*Luis Martín*) (*Álvarez*)  
 b. (*Luís*) (*Martín*) (*Álvarez*)

Deaccenting the first member of a complex first name is, in fact, frequent, while this never happens with complex surnames. The male first name *José* is pronounced with stress when it is the only first name, as in (27a), but in the complex first name *José María* the stress in the final syllable disappears and a secondary stress appears in the first syllable (*Jòse*) (27b).

- (27) a. *Jos[é] Pérez*  
 b. *J[ò]se Mar[í]a P[é]rez*

As can be seen in (27a), there is no avoidance of stress clash between the first name and the surname: the syllables /sé/ and /pé/ can both carry stress despite their being adjacent. The same is not true of complex first names. A colloquial form of the complex first name *José María* is *Josemari*, where compulsorily the syllable /se/ is deaccented, with rhythmic stress in the initial syllable. Similar stress-clash avoidance involves the male first name *Miguel* in combination with *Ángel*.

- (28) a. *J[ò]sem[á]ri*  
 b. \**Jos[è]m[á]ri*  
 c. *Migu[é]l*  
 d. *M[ì]guel [á]ngel*  
 e. \**Migu[é]l [á]ngel*

In complex surnames, there is no problem in having stress in two adjacent syllables; (29) contrasts with (28e).

(29) *Muñ[ó]z [á]lvarez*

This set of prosodic properties suggest that a combination of two first names is tighter – that is, they form a more cohesive constituent– than a combination of two surnames, or a combination between a first name and a surname: surnames form each its own intonation group, independent of other surnames and first names, so stress clashes might occur; first names form one intonation group, with a strong tendency to deaccent the first one, and avoiding stress clashes. If one assumes that prosody reflects syntactic structure (Wagner 2005) or that at least the prosodic structure is sensitive in part to syntactic labels (Nespor and Vogel 1986), this distinction is giving us information about two types of syntactic structures for first names and for surnames.

There are other, now syntactic, properties that differentiate combinations of first names and combinations of surnames. Even though related to a more formal style, surnames can be coordinated instead of juxtaposed; (30) is one grammatical way of expressing *Pedro Fernández López*.

(30) *Pedro [Fernández y López]*

In contrast, complex first names are never overtly coordinated. (31) is not a possible syntactic manifestation of the name *José María*, but is necessarily interpreted as the coordination of two referential expressions, one naming someone called José and one naming someone called María.

(31) #*José y María*

Again, this insists on the general idea that combinations of surnames are less cohesive than combinations of first names.

A contrast going in the same direction emerges when one considers contrastive focus structures. Imagine that we work in a company where there are two employees named María, María Fernández and María Pérez. If we need to contrast between them, we can do so with corrective negation and *sino* 'but'. Although less natural, negation could be external to the proper name (32b) and still scope over the surname.

- (32) a. *María, no Fernández sino Pérez*  
 María, not Fernández, but Pérez  
 b. *?No ha venido María Fernández, sino Pérez.*  
 not has come María Fernández, but Pérez

This is not possible with complex first names. Imagine that in this company we have two employees called María, but one is María José and the other is María Dolores, both with complex first names. We cannot contrast between the two of them as in (32).

- (33) a. *\*María, no José sino Dolores*  
 María, not José but Dolores  
 b. *#No ha venido María José, sino Dolores*  
 not has come María José, but Dolores

The only way of interpreting (33b) is that a woman called Dolores – not María Dolores – arrived, and (33a) can only be interpreted as somehow denying that a man called José – not someone called María José – is being referred to.

A final contrast is more lexical, but it also insists on the idea that the group formed by two first names is more cohesive than the one formed by two surnames: it is possible to have complex surnames consisting of two identical names (34a), but it is impossible to have two identical first names in a complex structure (34b).

- (34) a. *José Fernández Fernández*  
 b. *\*José José Fernández*

For this reason, a structure like (35) must necessarily be pronounced as (36a), with the middle element treated as a surname, and not like (36b), with the middle one treated as a first name.

(35) *Martín Martín Martín*

(36) a. *(Martín) (Martín) (Martín)*

b.\**(Martín Martín) (Martín)*

Let us now move to Belarusian patronymics and their properties.

#### 4. Belarusian patronymics

Remember that Slavic patronymics, here illustrated with Belarusian, are patronymics in a narrow sense that differentiates them from surnames and first names, as intermediate constituents which carry morphemes that are sometimes morphologically distinct from those that build surnames. In contrast to surnames in the same languages, the patronymic is semantically transparent –the first name of the father must be the base–. Patronymics are built with suffixes that always agree in gender with the referent, while only some suffixes used for surnames (*-ski*) show agreement (*-skaja*).

In what follows, instead of providing a full account of Belarusian patronymics, we will highlight the differences with Spanish surnames, beyond the agreement property.

First of all, it is generally not possible to use the patronymic alone in a proper name context. The example in (37c) contrasts with the examples in (37a) and (37b) in this regard. Native speakers consulted report that (37c) can be documented in some rural old-fashioned varieties, with a flavor of excessive colloquiality, but even there the feeling that one gets is that the patronymic is used as some sort of pet name to refer to the person, that is, instead of being a patronymic it is felt like some kind of alias or conventionalized way to refer to the person.



- (37) a. *Maryja Aliaksandrauna Sadoŭskaja*  
 Mary Alexander-PTR.FEM Sadouski.FEM  
*pryjechala ŭčora.*  
 arrive.PAST.3SG.FEM yesterday
- b. *Maryja Sadoŭskaja pryjechala ŭčora.*  
 c.% *Aliaksandrauna pryjechala ŭčora.*

This initial piece of data suggests that the patronymic should be viewed as an adjective, both due to its compulsory agreement with the first name and its inability to appear alone as a proper name, unless recategorized as its own proper name, a pet name of sorts.

Moreover, one can diagnose that the patronymic is an adjective that combines with the first name, not the surname. The following contrast suggests this: it is possible to have a sequence 'first name + patronymic', without the surname, but it is ungrammatical to have a sequence 'patronymic + surname' to the exclusion of the first name.

- (38) a. *Maryja Aliaksandrauna pryjechala ŭčora.*  
 b.\**Aliaksandrauna Sadoŭskaja pryjechala ŭčora.*

The constituency suggested by (38), with the first name and the patronymic forming a constituent to the exclusion of the surname, is reinforced by the prosody. The combination of the three parts of the proper name receive a prosodic packaging where the patronymic is with the first name, not the surname.

- (39) a. *(Maryja Aliaksandrauna) (Sadoŭskaja)*  
 b.\**(Maryja) (Aliaksandrauna Sadoŭskaja)*

Thus in parallel with Spanish, we would expect that the patronymic, like the elements of a complex proper name, should not allow for focalisation. However, this is not the case. With respect to the cohesion between the patronymic and the first name, we can diagnose that, like in the case of Spanish surnames, the patronymic and the first name show some inde-

pendence of each other. Imagine that in your company you have two women named Maryja, differentiated by the patronymic: Maryja Aliaksandrauna and Maryja Dzmitryeuna. Contrastive focus can be applied to the patronymics.

- (40) *Maryja, nie Aliaksandrauna a Dzmitryeuna.*  
 Mary not Aliaksandrauna but Dzmitryeuna

As expected, the same focalization can apply to surnames:

- (41) *Maryja, nie Sadoŭskaja a Caŭloŭskaja.*  
 Mary not Sadoŭskaja but Caŭloŭskaja

With complex first names, as in the case of Spanish, this type of focalisation is not allowed. Imagine that your company has a woman named Anna Maryja Aliena and one called Anna Sofya: the contrast in (42) is not possible: it would mean that some woman called Maryja is not the correct referent, but Sofya.

- (42) # *Anna, nie Maryja a Sofya.*  
 Anna not Maryja but Sofya

Thus, the properties of Slavic patronymics are somewhere in between combinations of first names and combinations of first names with surnames in Spanish, with the additional caveat that their distribution is similar to adjectives, rather than nouns. Along the same lines, the last relevant distinction emerges: patronymics cannot be grammatically iterated, even if one makes up the sufficient social conventions to grant that one person carries two patronymics (eg., one for the biological father, who died, and one for the adoptive father or the new husband of your mother, who died, or one for the mother and one for the father):

- (43) \* *Maryja Aliaksandrauna Dzmitryeuna Sadoŭskaja*

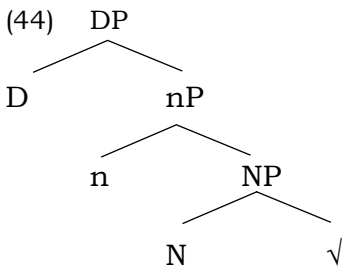
With these facts in mind, let us move to the proposed analysis.

## 5. Analyzing the internal structure of proper names

Let us start our analysis, first summarizing the generalizations identified:

- (i) Taken in the broad sense, patronymics exhibit in their morphological structure evidence of a segmentable morpheme that correlates with them being unable to function as first names; this property extends to its narrow sense, here illustrated with Belarusian.
- (ii) That patronymic in Spanish corresponds to the surname, which can appear alone as a proper name, and in Belarusian to an intermediate member of the full proper name that cannot appear naturally alone or with the surname in the absence of the first name.
- (iii) In Spanish, a complex first name forms its own prosodic phrase with a high degree of internal cohesion, generally involving deaccenting of the first name and avoiding stress clashes in general.
- (iv) There is an intonational break between the first name and the surname, which preserves stress in the two sides; in Belarusian the patronymic forms an intonational group with the first name.
- (v) Complex surnames assign a separate intonational break to each one of the surnames, and no stress clash leads to deaccenting of either surname.

Let us lay down our assumptions. We assume a complex functional structure for noun phrases projected above common nouns, including the following heads that are relevant for our analysis: D(eterminer), n (little n) and N (big N or lexical noun). These dominate a root that in the context of being the complement of N gets categorized as a lexical noun.

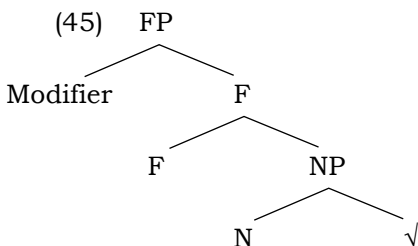


The role of D (Abney 1987) is to assign reference to the description provided by the noun. The determiner, by establishing reference and potentially identity if carrying the right type of head, has an anchoring function within the nominal domain that, following Wiltschko (2014), I take as equivalent to the role of tense in the sentential domain. As for N, which turns the root into a lexical noun, I follow Borer (2005) in the proposal that N is responsible for turning the root into a predicate of kinds – with the possibility that a further head turns it into a predicate of individuals given the right configuration. The root is assigned a conceptual meaning in the context of this categorizing head (Arad 2005), which associates with the constituent a set of properties which ultimately describe the type of kind or entity. N is, crucially, the head responsible for the descriptive content of a common noun. As for little *n*, I take it to be a functional nominal head lacking descriptive content, but responsible for several formal properties, most relevantly in Spanish or Belarusian the assignment of gender. Although not represented in the structure because they are not relevant for the analysis, I assume the standard heads for number and quantification, which are projected between D and *n*, as well as possible additional heads adding further descriptive content to the noun between *n* and N.

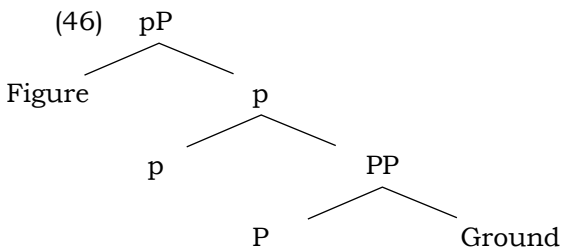
I assume, with strongly Neo-Constructionist approaches, that roots are elements deprived of syntactic features, including information about their grammatical category (Marantz 1997). The root, in this view, also lacks semantic information of its own, acting as nothing more than a placeholder to intro-

duce a specific lexical exponent (Borer 2013). Thus, the root in itself lacks syntactic and semantic properties, which are assigned to the set formed by the root and the first categorizer that dominates it.

The analysis will involve modification. I adhere to the idea that modification of a noun can be performed at several points in the structure (Svenonius 2008), depending on the type of content that is being modified. Thus, adjectives that intersect with the descriptive properties provided by N are introduced below n; those that provide information that is potentially relevant for the grammatical content of the nominal structure would be introduced below D. For explicitness, I assume Cinque's (2010) general take on modification as a specifier-complement relation intermediated by a functional head F, as in (45), which represents a modifier that intersects with the descriptive content of NP.



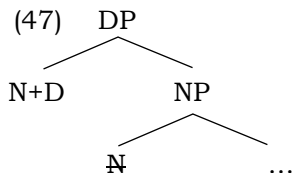
Finally, I will use relational heads –roughly corresponding to traditionally called 'adpositions' in my analysis of patronymics. I assume a structure for the prepositional area along the lines of (46), taken from Svenonius (2010).



With Talmy (2000) I assume that relational elements profile the relations between two entities in a figure and a ground, with the ground being taken as the reference point and the figure being the entity that is located in reference to it. Moreover, I take PP as the layer corresponding to the lexical content of the preposition: P names a relation adding conceptual semantics to it, with options such as Place, Path, Before, After, Instrument, etc., which given the right circumstances can combine with each other in complex PP structures. In contrast, the pP layer is merely functional and has the role of defining in syntax the relation by providing a position of the subject of which the relation is predicated, the figure. This means that functional prepositions without any associated content but required by syntactic conditions are projections of pP not involving PP. Prepositions with lexical content correspond to the structure of (46), and we will see that it is also possible to have a structure involving only P when the relation is lexically named but the second element is not syntactically defined.

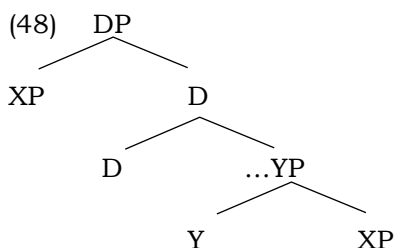
### 5.1. The syntax of proper names: basics

When it comes to the syntax of proper names, I mainly follow Longobardi (1994), but with a twist. In Longobardi's analysis, a proper name is inherently referential because it combines with DP, and carries some kind of feature that triggers movement of N (in his analysis) to the head D, as represented in (47) – with Longobardi's labels.

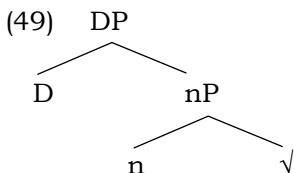


Assuming the core of the proposal, I introduce two minimal changes:

- (i) Even if head movement can explain most of the facts for simple proper names like *Mary*, it does not explain the fact that proper names can be complex, that is, more than heads. I assume, therefore, that when the structure is more complex below DP, the proper name can still satisfy its referentiality by phrasal movement to spec, DP. In such cases, D remains silent, I assume, because of Koopman's (2000) Generalized Doubly-Filled Comp filter, which precludes double spell out of a specifier and a head that share features – in our case, referentiality.



- (ii) I adhere to the non-descriptive theory of proper names where these are taken to be rigid designators (Kripke 1980). This, in my view, has the consequence that NP is not projected within the structure of a proper name, because at no point is the proper name a predicate characterized by a set of properties. However, proper names carry gender, which means that nP is projected in their structure (see also Fábregas 2020 for further arguments of this). Thus, my proposal is that, while (44) corresponds to the structure of a common noun, (49) corresponds to that of a proper name.

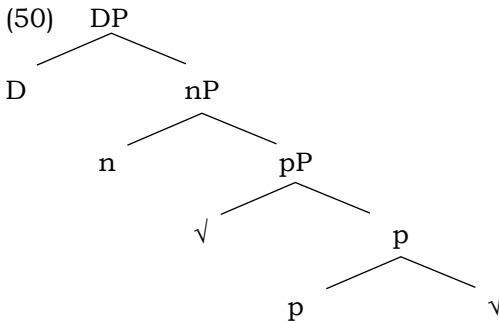


Lacking descriptive properties, proper names cannot be modified by adjectives and still have the grammatical distribution of

proper names. The way in which this theory explains the fact that proper names combined with adjectives must combine with a determiner is that in those contexts they are projected as common nouns (thus, the structure of 44) because the adjective needs NP to be projected. Unlike Longobardi, I do not explain the facts in (21)-(23) as the effect of the adjectival head, or the F head that introduces it, blocking head movement of the proper noun to D, in essence because, as I just said, I assume phrasal movement when the proper name is complex.

## 5.2 Complex first names

The tree in (49) corresponds to a simple first name, like *José*. For the case of complex first names, we propose the following structure: two roots that are embedded under one single nP layer, which nominalizes both of them at the same time.



The structure that we have used to express the relation between the two roots is a functional relational structure, little p (Svenonius 2010). Remember that Spanish first names cannot be combined with each other through a copula, unlike surnames (cf. 31 above). Moreover, someone that is called *Juan José* is not someone that is called Juan and is also called José, as a coordination, but rather someone that carries this complex name. Embedding the two roots through a relational projection that combines them both, and making that set be further nominalized by little n, is the device that we use pre-



cisely to express this lack of coordination and the fact that the complex first name acts as a single unit syntactically.

My claim, given that structure, is moreover that the real combinations of first names is restricted to two, because pP is not iterable. Social conventions allow in Spanish that people receive more than two first names legally, creating combinations such as the well-known (for speakers from Spain) *Felipe Juan Froilán de todos los Santos*. However, these are not used in the language, where such long complex first names are always reduced to one or at most two first names – in our case, *Froilán*.

(50) expresses also the property identified in (33) that the two members of the complex first name cannot be separated by contrastive focal scope. For the purposes of the structure, there is only one nP, which is taking as its root a complex structure where two –or potentially more roots– are contained. Remember that the root lacks its own semantic and syntactic properties. Lacking a grammatical category and other syntactic and semantic properties, each root cannot be affected independently by corrective negation because the negative operator lacks the syntactic or semantic information to act over the root.

(51) \**[María]<sup>√</sup>, no [José]<sup>√</sup> sino [Luisa]<sup>√</sup>*

Negation can affect, however, an nP, which in this case dominates the two roots, because little n adds syntactic and semantic features to the representation which negation can operate over.

(52) *no [María José]<sup>nP</sup> sino [María Luisa]<sup>nP</sup>*

The prosodic properties are also explained by this structure, in a simple way: the constituent that is minimally categorized in (50) is the complex formed by the two roots. Both of them, as a complex constituent, get assigned the nominal category by the same head, little n. Crucially, it is not the case that each root is dominated by its own nP.

We can assume that this is the minimal size of a prosodic constituent in Spanish, as in fact has been argued in the literature: the first categorizing head in the sequence, in our example little *n*, defines its domain as a prosodic word (Arad 2005, Bermúdez Otero 2013). Given that the two roots are categorized together by *nP*, the two roots form one single prosodic constituent, which can maximally carry one stress. This explains the deaccenting of the first element and the stress clashes that trigger changes in the tonic syllable of the first name.

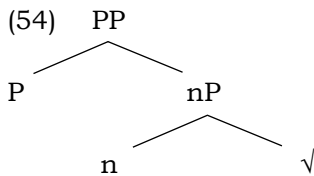
(53) (*Jòse María*)<sup>w</sup>

### 5.3. Spanish surnames

Let us now move to the Spanish surname, specifically the morphological patronymics. We have seen evidence that the segment *-ez* corresponds to a segmentable constituent. Its semantics, as we have seen, is relational, as evidenced by the cases that are semantically compositional ('son or daughter of BASE'). Moreover, patronymics can be built, cross-linguistically, through prepositions, as we saw for Portuguese above.

All these properties lead me to propose that the layer corresponding to the suffix *-ez* in Spanish is a manifestation of a PP layer with lexical content, where the conceptual meaning related to the relation is the one corresponding to 'child of'. In the compositional cases, as it will be the case of the Belarusian patronymic, this is unproblematic because it directly reflects in the structure. In the non-compositional cases – like Spanish surnames, where being called *Fernández* identifies which family you belong to but does not let you infer that the father is called *Fernando* – I still claim that the relation of being the descent of someone else stays, although deprived of the relational content. That is: the surname is marking that one belongs to the same clan or family as someone else, without expressing overtly the relation between the father or the mother. This is

precisely what it means to be a projection of PP without pP: there is a relation, but it is named through the conceptual content, without the syntactic structure providing support to introduce the second member of the relation. Thus, (54) is the structure of a surname in Spanish.<sup>2</sup>



The base of the noun is itself categorized by little *n*, as a proper name. Remember that *-ez* does not displace the stress of the first name (12 above). This theory explains why: the *nP* layer, as in the case of complex first names, defines a minimal prosodic constituent where the stress is assigned. The patronymic suffix is outside that domain, so once it is added to the word, it cannot modify the prosody of the *nP* layer.

At this point, I would like to say something about surnames that, being surnames with all grammatical properties of them, lack any overt patronymic suffix, such as those in (55).

(55) *Marín, Arche, Vivanco, Gibert, Acedo, Fábregas...*

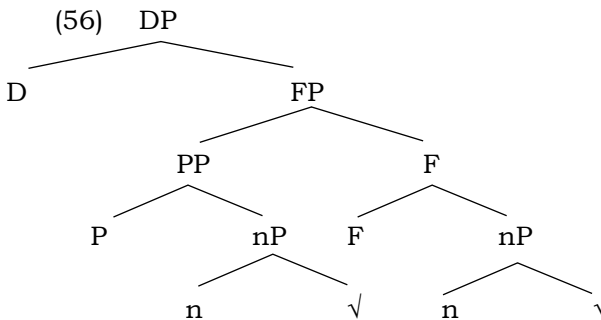
My claim is that these surnames also correspond to the structure in (54), only that in their case the exponent used to spell out the structure includes, as a portmanteau morpheme or a synthetic morpheme, also the head corresponding to *P*. Sur-

---

<sup>2</sup> An anonymous reviewer, whom I am very grateful to, notes that perhaps this can be related to the fact that Spanish surnames lack a plural form, in contrast to first names (cf. *los Martín-es* 'the Martin-s', which is grammatical if *Martín* corresponds to a first name but ungrammatical if it corresponds to a surname). I find this idea worth pursuing: PPs lack plural forms. From this approach, languages where surnames have plural forms are either agreeing surname languages – where the plural is actually agreement with a noun – or nominal structures not including *P*, and their properties should be radically different from Spanish surnames. A typological survey might confirm or reject this initial hypothesis.

names in *-ez* let us see the syntactic relation between the base and the patronymic, so that we can identify a structure that is identical for all surnames, including those that spell the PP layer together with the root exponent.

(56) presents my proposal about how the surname can be combined with the proper name. I take it as a modifier of the first name that is merged, like all modifiers, in a specifier position.



Remember that, with the rigid designator theory, I take proper names to lack descriptive properties, so this modification does not result in anything like set intersection. The surname, however, like other modifiers, restricts the modified element: in our case, it restricts the reference of the first name to those that also carry the specific surname, so that different Mariás can be differentiated in the appropriate context.

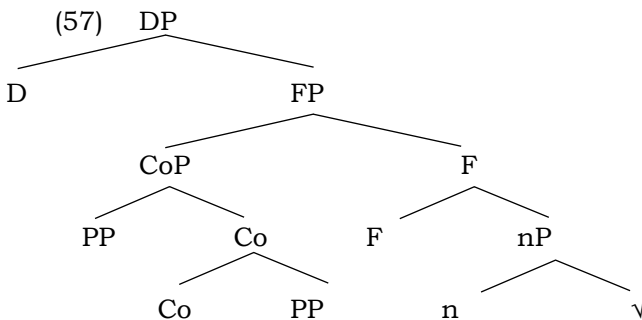
Assuming head movement of the root+n to D or phrasal movement of nP to spec, DP would result in the right syntactic order: first name + surname. I speculate that languages where the convention is to place the surname before the first name, like Hungarian, simply reflect the base generated order without movement to D or DP.

This structure captures the property that the surname can be contrastively negated without the first name and that it establishes its own prosodic constituent, independent of the first name. Note that in the structure the surname constitutes a categorized complex specifier, that is, a specifier consisting

of a projection of XP and not a simple head, as it was the case with the roots in the complex first name. Of course, we know that complex specifiers behave as islands for syntactic extraction, and as units from a phonological and even semantic perspective. Uriagereka's (1999) Multiple Spell Out theory has proposed that complex specifier always behave as closed domains for prosody because, in essence, they must undergo spell out before they are introduced in the derivation and merged as specifiers as another category in the spine of the tree.

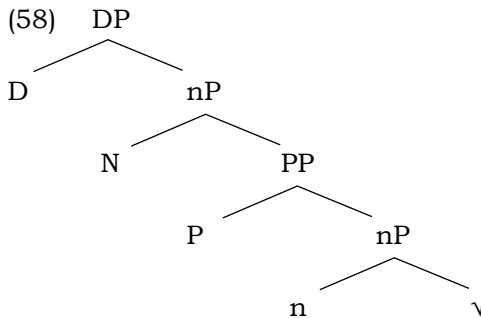
This, in itself, is enough to account for why the surname and the first name belong to two different prosodic domains. It also accounts for the fact that the surname can be negated contrastively because, unlike each member of the complex first name, it is a categorized element that contains syntactic and semantic information.

Surnames can also be complex, and in fact in Spanish they can be overtly coordinated. For them, I simply propose that the modifier in their case is a coordinated phrase –where coordination can be phonologically silent or overt–, as in (57).



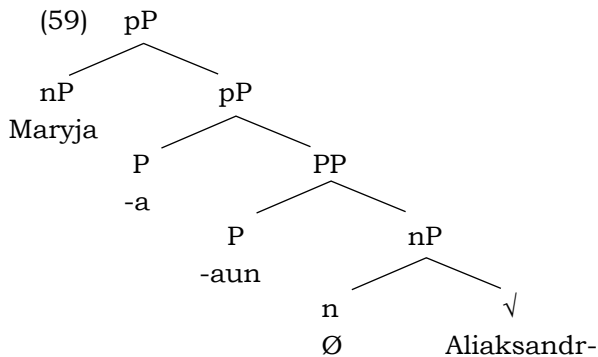
Note that within the coordinated structure (CoP), the first surname is also a complex specifier internal to the coordination; this guarantees, by Multiple Spell Out, that each surname in the coordination will correspond to its own prosodic phrase: each one carries its own stress and stress clash is not avoided.

Finally, let me briefly address what I assume to be the structure of a surname used without first name, as a proper name. In such cases I assume that the surname has been further nominalized, as in (58). As without the first name the surname does not have anything to modify, head movement is possible because all heads are in a sequence with D.



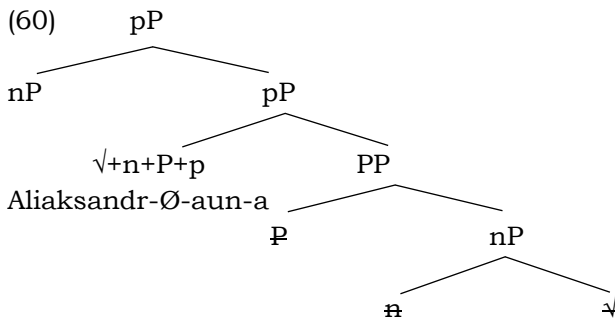
#### 5.4 Belarusian patronymics

Remember that Belarusian patronymics have two internal differences with surnames in Spanish: they are sensitive to gender agreement, like adjectives, and they are compositional in that they directly express the relation 'son / daughter of BASE', where the base is the first name of the father. I take both differences to follow from the patronymic carrying not only a PP layer, but also a pP layer, as in (59), where I represent the structure of *Maryja Aliaksandrauna*.



The syntactic expression, containing a full prepositional structure, directly reflects the interpretation 'María, daughter of Aleksander'. Importantly, this structure also explains why the patronymic must be with the first name and not with the surname: they belong to the same syntactic constituent, under pP. Note, moreover, that (59) explains why the patronymic can be contrasted excluding the first name: in this structure, the segment corresponding to the patronymic is not simply a root, but a categorized constituent, like a Spanish surname and not like one of the two members of a complex first name.

Assuming head movement of the root to p, as in (60), the structure also explains why –even if the patronymic can be negated independently of the first name– they end up being within the same prosodic domain: after head movement in order to get the patronymic suffix and the gender marker, the root ends in p, while the first name is in spec, pP, therefore obligatorily ending adjacent to each other.



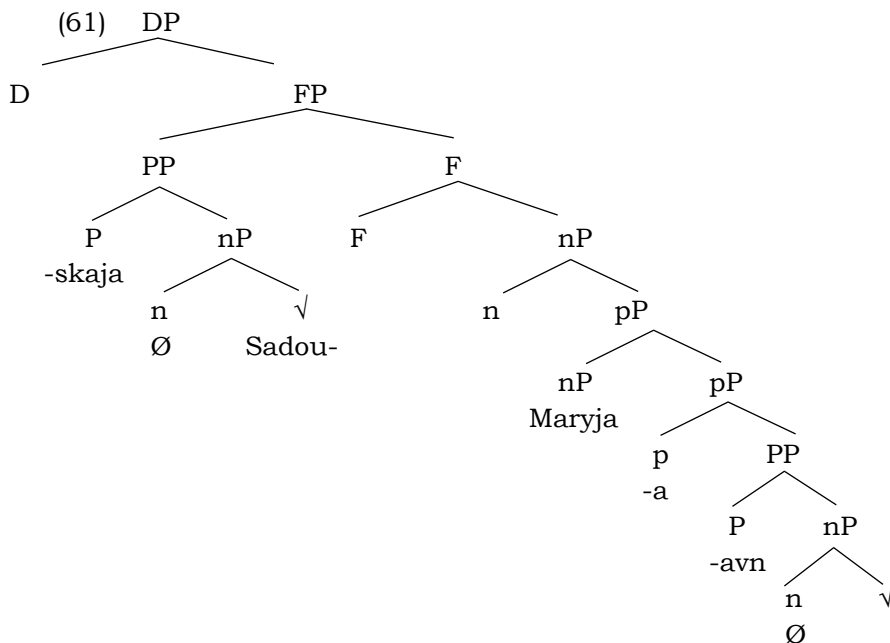
The proposed structure explains why the Belarusian patronymic cannot be iterated and why it cannot stand alone as a proper name, unlike the surname in Spanish: the pP structure syntactically defines the relation as biunivocal, setting only one ground and only one figure by virtue of the presence of pP. By the same reason, pP forces the patronymic taken as its complement to act necessarily as one of the two members of a relation, so that it cannot appear in the absence of the first name.

The feeling that the patronymic used alone acts as a 'pet name', in this view, is explained in the following way: instead of being used as a patronymic with the structure in (59), in those cases the patronymic projects only up to PP, which is further nominalized as in the case of the Spanish surname (58).

In section §2 above we suggested a possible generalization tying together agreement and compositionality in the interpretation of the patronymic: if you are a semantically compositional patronymic, you must agree in gender. Our structure gives a chance to explain this generalization. In (59), the compositional interpretation is related to the presence of pP, which defines at a syntactic level the relation named by the PP layer. The head p, however, is a functional head that provides syntactic structure, not lexical meaning. My proposal is that gender agreement is introduced in compositional patronymics by the pP layer, which at the same time provides the compositional reading and the features that trigger agreement in gender with its specifier, the first name Marya, which is feminine. However, we insist that the small set of languages considered do not grant that the generalization is correct. Moreover, the generalization is not biconditional, as surnames in Belarusian can be sensitive to gender agreement even though they are not compositional, suggesting that P could in principle also host gender features.

Beyond this, for Belarusian surnames we assume the same type of structure as in Spanish. (61) represents the whole structure; we assume movement of the higher nP to spec, DP in Belarusian to obtain the right order between the components of the proper name.





## 6. Conclusions

In this article we have researched the internal syntax of proper names, triggered by the existence of a class of derived patronymics, both in Spanish and Belarusian, which can be morphologically decomposed. Our proposal highlights several properties that could be eventually checked in further research:

- (i) Proper names have a complex syntactic structure.
- (ii) Relational structure, PP, pP or both, is involved in relating the components of proper names together: surnames are modifiers that restrict the reference of a proper name, and patronymics are grounds which profile first names as figures.
- (iii) As a preliminary hypothesis, the compositional reading of patronymics involves projection of a functional relational structure which must contain agreement features that make the patronymic sensitive to the gender of the first name.

The overarching conclusion of this paper is twofold:

- (a) morphological elements can be used as triggers to identify complex internal structures, even in a domain like proper names where one initially assumes a quite rigid functional structure;
- (b) morphological decomposition can be performed also in cases where there is no compositionality in meaning, because the morphemes involved define different types of configurations and formal properties that are significant for syntax, even when they are not so directly translatable for semantics.

## References

- Abbott, Barbara (2002). "Definites and proper names: Some bad news for the description theory". *Journal of Semantics* 19: 203-207.
- Arad, Maya (2005). *Roots and patterns*. Dordrecht: Springer.
- Bermúdez Otero, Ricardo (2013). "The Spanish lexicon stores stems with theme vowels, not roots with inflectional class features". *Probus* 25: 3-103.
- Evans, Gareth (1982). *Varieties of Reference*. Oxford: Oxford University Press.
- Fábregas, Antonio (2020). *Morphologically Derived Adjectives in Spanish*. Amsterdam: John Benjamins.
- Kaplan, David (1989). "Demonstratives". In: Joseph Almog, John Perry, Howard Wettstein (eds.). *Themes from Kaplan*. Oxford: Oxford University Press, 481-563.
- Kohlheim, Rosa, Volker Kohlheim (2000). *Duden. Lexikon der Vornamen: Herkunft, Bedeutung und Gebrauch von über 8000 Vornamen*. Mannheim – Zürich: Dudenverlag.
- Koopman, Hilda (2000). *The Syntax of Specifiers and Heads*. Oxford: Routledge.
- Kripke, Saul (1980). *Naming and Necessity*. Cambridge, Mass.: Harvard University Press.
- Longobardi, Giuseppe (1994). "Reference and proper names: A theory of N-movement in syntax and logical form". *Linguistic Inquiry* 25: 609-665.

- Matushansky, Ora (2008). "On the linguistic complexity of proper names". *Linguistics and Philosophy* 31: 573-627.
- Nespor, Marina, Irene Vogel (1986). *Prosodic Phonology*. Dordrecht: Foris.
- Recanati, François (1997). *Direct Reference: From Language to Thought*. Oxford: Blackwell.
- Sainsbury, Mark (2015). "The same name". *Erkenntnis* 80: 195-214.
- Svenonius, Peter (2010). "Spatial P in English". In: Guglielmo Cinque, Luigi Rizzi (eds.). *Mapping Spatial PPs: The Cartography of Syntactic Structures*. Vol. 6. Oxford: Oxford University Press, 127-160.
- Talmy, Leonard (2000). *Toward a Cognitive Semantics*. Cambridge, Mass.: MIT Press.
- Uriagereka, Juan (1999). "Multiple spell-out". In Samuel D. Epstein, Norbert Hornstein (eds.). *Working Minimalism*. Cambridge, Mass.: MIT Press, 251-281.
- Van Langendonck, Willy (2007). *Theory and Typology of Proper Names*. Berlin – New York: Mouton de Gruyter.
- Wagner, Michael (2005). *Prosody and Recursion*. PhD dissertation, MIT.

Antonio Fábregas  
ORCID iD: 0000-0001-9907-5878  
University of Tromsø  
The Arctic University of Norway  
Hansine Hansens veg  
9037 Tromsø  
Norway  
antonio.fabregas@uit.no



## **Non-verbal complements of modal verbs: The case of directional adverbs in Czech**

HANA GRUET-SKRABALOVA

*Received 03.05.2021,  
received in revised form 10.11.2021,  
accepted 15.11.2021.*

### **Abstract**

This paper deals with structures where Czech modal verbs (*muset* 'must', *moci* 'can', *smět* 'be allowed') combine, at surface, with an adverbial complement and which involve an event of movement to the place denoted by this complement. Since modal verbs normally select a VP complement, the question arises whether these structures contain an elided or a null verb GO, or whether modal verbs here directly select a directional adverbial, whose motion interpretation supplies a 'missing' verb of movement. We show in this paper that there is not enough evidence to posit a null lexical verb GO in the structures under discussion. We then argue that these structures are licensed by modality like non-finite or non-sentential wh-clauses that may also contain a directional adverbial without an overt verb of movement. However, in declarative clauses, which require a verbal head to bear tense and agreement feature and to support the negative prefix *ne-* expressing sentential negation, the modality must be overtly realized by a modal verb.

### **Keywords**

modal verbs, directional adverbs, verbs of movement, Czech, Slovenian

## **Dopełnienia domyślne czasowników modalnych: przypadek przysłówków kierunku w języku czeskim**

### **Abstrakt**

Niniejszy artykuł dotyczy struktur, w których czeskie czasowniki modalne (*muset* „musi”, *moci* „można”, *smět* „być dozwolonym”) łączą się, *prima facie*, z okolicznikiem, przez co wyrażają ruch do miejsca wyznaczonego ten okolicznik. Ponieważ czasowniki modalne zwykle wybierają dopełnienie VP, pojawia się pytanie, czy mamy do czynienia z elipsą czasownika bądź formą zerową czasownika GO, czy też czasowniki modalne w tym przypadku bezpośrednio wybierają przysłówki kierunku, którego interpretacja ruchu dostarcza „brakującego” czasownika ruchu. W tym artykule pokazujemy, że nie ma wystarczających dowodów, aby zakładać istnienie zerowej formy czasownika leksykalnego GO w omawianych strukturach. Następnie argumentujemy, że te struktury są licencjonowane przez zdania składowe z czasownikami w formie osobowej lub nieosobowej, które mogą również zawierać przysłówki kierunku bez wyrażonego czasownika ruchu. Jednak w zdaniach oznajmujących, które wymagają, aby rdzeń frazy czasownikowej zawierał cechę czasu i zgodności oraz wspierał przedrostek przeczący *ne* – wyrażający negację zdaniową, czasownik modalny musi być wyrażony.

### **Słowa kluczowe**

czasowniki modalne, przysłówki kierunku, czasowniki ruchu, język czeski, język słoweński

### **1. Introduction**

This paper deals with structures where modal verbs in Czech combine, at surface, with an adverbial complement (mostly PP) like in (1) and which involve an event of movement to the place denoted by this complement. We focus on strict modals *muset* (must), *moci* (can) and *smět* (be allowed), in comparison with

the volitional verb *chtít* (want),<sup>1</sup> which allows for the same directional complements, see (2).<sup>2</sup>

## (1) Czech

- a. *Musím do Prahy/ k doktorovi.*  
must.1SG to Prague.GEN to doctor.DAT  
'I have to go to Prague'/ 'I have to go to the doctor.'
- b. *Po tom infarktu může jenom na zahradu.*  
after that heart attack can.3SG only on garden.ACC  
'After his heart attack, he can only go out in the garden.'
- c. *Po tom infarktu nesmí ani na zahradu.*  
after that heart attack NEG.can.3SG even on garden.ACC  
'After his heart attack, he is not allowed to even go out in the garden.'

## (2) Czech

- a. *Chceš do kina?*  
want.2SG to cinema.GEN  
'Do you want to go to the cinema?'
- b. *Chce se ti do kina?*  
want.3SG.NEUT REFL you.DAT to movie.GEN  
'Do you feel like going to the cinema?'

In contrast, this structure is impossible with other lexical verbs that may combine at surface with motion verbs and their directional complements, like *zkusit* 'to try', *akceptovat* 'to accept', *rozhodnout* 'to decide':<sup>3</sup>

<sup>1</sup> When relevant, the differences between the strict modals and the verb *want* are explicitly put forward in the paper.

<sup>2</sup> Abbreviations in glosses: ACC: accusative case, DAT: dative case, GEN: genitive case, IMP: imperfective, PF: perfective, FUT: future, REFL: reflexive, SG: singular, PL: plural, NEUT: neuter, NEG: negation, POSS: possessive.

<sup>3</sup> These verbs do not allow for NCA in Czech:

- (i) *Zítřa zkusím jít do kina. / Zítřa \*(to) zkusím.*  
tomorrow try.1SG go to cinema.GEN tomorrow it try.1SG  
'Tomorrow I'll try to go to the cinema./Tomorrow I'll try (it).'
- (ii) *Akceptoval jít dovnitř. / Akceptoval \*(to).*  
accepted go inside accepted it  
'He accepted to go inside.' / 'He accepted (it).'

- (3) Czech
- a. *Zítřa zkusím jít do kina.* / \**Zítřa zkusím*  
 tomorrow try.1SG go to cinema.GEN tomorrow try.1SG  
*kina.*  
 cinema.GEN  
 ‘Tomorrow I’ll try to go to the cinema’
- b. *Nakonec akceptoval vejít dovnitř.* / \**Akceptoval dovnitř.*  
 finally accepted in.go inside accepted inside  
 ‘Finally, he accepted to go inside.’

The structures in (1) and (2) raise thus a question with respect to the selectional properties of modal verbs: since modal verbs normally select a VP complement, should we assume that the structures in (1) and (2) contain an elided or a null verb GO,<sup>4</sup> or some null copula verb? Or should we better account for these structure by assuming that modal verbs may directly select a directional adverbial, whose motion interpretation supplies a ‘missing’ verb of movement?

Though we are not against the idea of a null verb GO in the grammar, we show in this paper that there is not enough evidence to posit such a null lexical verb in structures with strict modal verbs and directional adverbials in Czech. Rather, we argue that these structures are licensed by modality like non-finite or non-sentential wh-clauses. However, in declarative clauses, which require a verbal head to bear tense and agreement feature and to support the negative prefix *ne-* expressing sentential negation, the modality must be overtly realized by a modal verb.

The paper is organized as follows. In section 2, we show basic arguments against ellipsis analysis. In section 3, we present arguments for a null verb GO that have been put forward by van Riemsdijk for Germanic languages and by Marušič and Žaucer (2005) for Slovenian. We then argue that there is no straightforward evidence for positing a null verb GO in

---

<sup>4</sup> GO means an abstract verb that can be lexicalized by *jít* (go) or *jet* (ride) or their imperfective or perfective variants.



Czech. In section 4, we propose an alternative GO-less analysis of our structure.

## 2. Against an ellipsis account

Modal verbs in Czech allow for ellipsis of their VP-complement, as we can see in (4a). Structures involving ellipsis differ however from structures involving adverbials in at least two aspects. First, elided elements need a linguistic antecedent<sup>5</sup> in order to be licensed and interpreted, compare (4a) with the VP-antecedent and (4b) without antecedent.

- (4) a. *Jan si mohl vzít dovolenou v srpnu, ale jeho*  
 Jan REFL could take vacation in August but his  
*kolega nemohl.*  
 colleague NEG.could  
 ‘Jan was allowed to take his vacation in August but his colleague was not.’
- b. \* *Janův kolega nemohl.*<sup>6</sup>  
 Jan.POSS colleague NEG.could  
 ‘John’s colleague could not.’

Second, VP-ellipsis requires a contrastive remnant element. Typically, this contrastive element will be the subject, as in (4), but it can also be the complement of the non-finite lexical verb, that had been topicalized and moved out of the VP before ellipsis took place, as shown in (5). In the same way, the contrast may involve adverbial elements, like in (6). Note that in absence of the VP antecedent, the sequence ‘modal verb + locative adverbial’ in (6b) is ungrammatical.

---

<sup>5</sup> Exophoric antecedents of VP-ellipsis in Czech are limited to conventional cases (Gruet-Skrabalova 2016), and even in such cases, we can assume they contain a null verbal anaphor rather than an elided VP. This verbal anaphor gets its interpretation from the situation that is going on:

- (i) *Ja už nemůžu.*  
 I yet NEG.can.1SG  
 ‘I cannot do what I am doing anymore.’

<sup>6</sup> Example (4b), (5b) and (6b) are unacceptable without a previous context.

- (5) a. *Latinu můžu učít já a řečtinu může*  
 Latin.ACC can.1SG teach I and Greek.ACC can.3SG  
*můj kolega*  
 my colleague  
 'I can teach Latin and my colleague can teach Greek.'
- b. \* *Řečtinu může můj kolega.*  
 Greek.ACC can.3SG my colleague  
 '\*My colleague can Greek.'
- (6) a. *Doma si hrát můžete ale venku nesmíte.*  
 home.LOC REFL play can.2PL but outside.locNEG.can.2PL  
 'You may play at home, but you can't outside.'
- b. \* *Venku nesmíte.*  
 outside.loc NEG.can.2PL  
 '\*Outside you can't.'

Importantly, no antecedent is necessary to license and interpret the sentences in (1) and (2)<sup>7</sup> above. Likewise, the directional PP is not necessarily contrastively focused with another PP in the context. We conclude thus that ellipsis of a lexical verb of movement can be thus be easily rejected. In the next section, we turn to another analysis that considers that the structures we are dealing with contain a null verb GO.

### 3. Arguments in favour of a null verb GO

Structures like in (1) exist in other languages. Van Riemsdijk (2002) argues for a null verb GO in Germanic languages (except for English) like in Swiss German in (7a). His main argument comes from the contrast between (7a) and (7b) with respect to the position of the adverbial element *hăi* (home). When a lexical verb of motion is present in the sentence, the adverbial-

<sup>7</sup> Gruet-Skrabalova (2019) also shows that even epistemic modal verbs allow for ellipsis:

- (i) *Může to být pravda, ale nemusí.*  
 can it be true but NEG.must  
 'It might be true, but it doesn't have to.'

In contrast, modal verbs combining with directional adverbials have always deontic reading, see section 3.1.

al cannot occur in clause-final position, see (7b). The fact the adverbial in (7a) is acceptable in clause-final surface position suggests that it is followed by a null verb GO. Since Czech is not a V2 language, this argument cannot be applied to our data.

(7) Swiss German

- a. ... *wil mer (häi) hetted (häi) sole (häi).*  
 because we home would home had to home  
 ‘...because we should’ve gone home.’
- b. ... *wil si iri tochter (häi) hetted (häi)*  
 because theytheir daughter home would home  
*sole (häi) schicke (\*häi)*  
 had to home send home  
 ‘...because they should’ve sent their daughter home.’

Marušič and Žaucer (2005) argue for the existence of a null verb GO in Slovenian, whose distribution would be however larger than that of a null GO in Germanic. In the next subsections, we discuss the main arguments they present in favor of their claim: the presence of contradictory temporal adverbials, the use of purpose PPs, VP conjunction and covert modality. We argue that these arguments are not really conclusive, at least for Czech.

### 3.1. Temporal adverbials

Marušič and Žaucer (2005) argue that the possibility to have two contradictory temporal adverbs in (8b), but not in (8a), indicates that the sentence (8b) contains two temporally independent events and thus a syntactic structure with two VPs. The second adverb in (8b) would thus be dependent of the VP involving the null verb GO.

## (8) Slovenian

- a. \* *Včeraj Lina ni šla jutri domov*  
 yesterday Lina not go.PST tomorrow home  
 ‘Yesterday, Lina didn’t go home tomorrow.’
- b. *Včeraj se Lini ni ljubilo jutri domov.*  
 yesterday NON-ACT Lina.DAT not feel.PST tomorrow  
 home

In Czech, the simultaneous presence of ‘yesterday’ and ‘today’ in (9) is infelicitous even when the modal verb is followed by the overt verb *jít* ‘go’. Note however that the verb in (9a) has deontic reading (*i.e.* he had to go to the doctor at a moment *x*). The example (10a), where the modal verb has epistemic reading (*i.e.* he thought it necessary (yesterday) to go to the doctor (today)) is acceptable. It is thus the obligation of ‘going somewhere’ that cannot be situated at another moment that the event of ‘going somewhere’ itself. The fact that the verb ‘go’ cannot be omitted in both (9b) and (10b) implies that the structures where modals combine with a directional PP only have deontic reading. This is actually the case in (1) and (2) above.

## (9) Czech

- a. \* *Včera musel jít k doktorovi dneska.*  
 yesterday must.PR.3SG.M go to doctor.DAT today  
 ‘(Intended:) Yesterday he had to go to the doctor today.’
- b. \* *Včera musel k doktorovi dneska.*

## (10) Czech

- a. *Včera musel jít k doktorovi už dneska (a dneska říká, že už ho to nebolí).*  
 yesterday must.PR.3SG.M go to doctor.DAT already  
 today and today say.3sg that already he it  
 NEG.ache.3SG
- b. \* *Včera musel k doktorovi už dneska.*

In contrast, with the volitional verb *chtít* (want), the contradictory temporal adverbials are acceptable both with and without the overt verb:<sup>8</sup>

(11) Czech

- a. *Včera chtěl odjet na dovolenou už  
yesterday wanted PF.go on holidays already  
příští týden, a dneska chce odjet až za dva  
next week but today want PF.go only in two  
týdny.*

Weeks

'Yesterday, he wanted to leave on holiday already next week, and today he wants to leave only in two weeks.'

- b. *Včera chtěl na dovolenou už příští týden, a dneska chce až  
za dva týdny.*

Since the structures with strict modal verbs and directional adverbials only involve one event, we conclude that they do not require the presence of a null verb GO. Even if we admit that a single node T could be compatible with two V nodes,<sup>9</sup> we consider that the data like in (9) and (10) do not establish a strong piece of evidence in favor of a null V head.

### 3.2. Purpose PPs

A modal verb in Slovenian can combine not only with a directional PP, like in (12a), but also with a non-directional PP with the 'purpose' preposition *po*, like in (12b), which implies 'movement with a purpose'. Since *po* cannot occur with other than motion verbs, Marušič and Žaucer (2005) claim that in (12b),

<sup>8</sup> This is also true for the verb *chtít* when it is used in neuter form with a dative subject, meaning 'feel like':

(i) *Včera se mu zachtělo (jít) domů už dneska  
yesterday REFL he.DAT PF.wanted go house already today  
'Yesterday he felt like going home today.'*

<sup>9</sup> This has been noted us by an anonymous reviewer. I am thankful to this reviewer for his/her constructive remarks.

which is read as ‘he must go and get bread’, a motion verb must actually present but is not pronounced.

- (12) Slovenian
- a. *Vsak Slovenec mora vsaj enkrat na Triglav.*  
 every Slovenian must at-least once onto Triglav  
 ‘Every Slovenian must go up Mt. Triglav at least once.’
- b. *Peter mora (v trgovino) po kruh.*  
 Peter must to store for bread  
 ‘Peter must go (to the store) and get some bread.’

In Czech, the PP after the modal verb can also be introduced by the purpose preposition *pro* (for) or *na* (on/for), as shown in (13). Such a purpose PP is not by itself directional but it implies a place where we have to go in order to get the DP introduced by *pro* or *na*. This place can be stated explicitly by a directional PP, as shown in (14). But usually, the directional PP is not necessary, because it can be inferred from the purpose PP itself: the croissants are bought in a store, the mushrooms grow in the forest, and the children have to be picked up from school.

- (13) Czech
- a. *Zeptej se ho jestli může pro pár*  
 ask REFL he.ACC if must.3SG for some  
*rohlíků.*  
 croissants.GEN  
 ‘Ask him if he can go and get some croissants.’
- b. *Už jsou čtyři, musím pro Adama.*  
 already are four must.1SG for Adam.ACC  
 ‘It’s 4 o’clock, I have to go and pick up Adam.’
- c. *Dneska odpoledne můžem třeba na houby.*  
 today afternoon can.1PL maybe for mushrooms  
 ‘This afternoon we may go and look for mushrooms.’
- (14) Czech
- a. *Zeptej se ho jestli může do obchodu (=to the store) pro pár rohlíků.*

- b. *Už jsou čtyři, musím do školky* (=to the kindergarden)  
*pro Adama.*
- c. *Dneska odpoledne můžem třeba do lesa* (=to the  
 forest) *na houby.*

We claim thus that a purpose PP do not require an overt verb of movement. Rather, we consider that an event of movement implies both direction and goal (see section 4 for more details), which has as consequence that purpose PPs appear in the same modal structures that directional PPs. We conclude that the data discussed here do not necessarily imply the presence of a null verb GO in structures where modal verbs combine with a purpose PP.

### 3.3. Coordination

In Slovenian, modal verbs can have scope over conjunction. Assuming that conjuncts must be identical, Marušič and Žaucer (2005) claim that both conjunct in examples like (15) must be (at least) VPs.

- (15) Slovenian  
*Vid ni mogelveč niti do avta niti postaviti*  
 Vid not could still neither to car neither put-up  
*šotoru.*  
 tent  
 'Vid could neither go to the car nor put up a tent.'

However, that there has been shown in the literature (e.g. Bayer 1996) that categorial identity of conjuncts is not obligatory. The conjuncts need to be semantically compatible and able to appear alone in the position of the coordinate phrase, as we can see in (16). It is thus not surprising that we can conjoin directional adverbials after modals with an overt VP, as in (17), since both may function as a predicative phrase. These examples can simply be analyzed as involving two conjoined PredPs.

- (16) Czech
- a. *Doktor mu doporučuje víc klidu a nekouřit.*  
 doctor he.DAT recommend.3SG more rest.GEN and  
*nekouřit.*  
 NEG.smoke  
 ‘The doctor recommends that he rests more and stops smoking.’
- b. *Doktor mu doporučuje víc klidu / nekouřit.*  
 ‘The doctor recommends that he rests more/that he stops smoking.’
- (17) Czech
- a. *Dneska odpoledne můžem třeba do lesa a zaplavat si.*  
 today afternoon can.1PL maybe to forest.GEN and  
*zaplavat si.*  
 PF.swim REFL  
 ‘Today afternoon we may have a walk to the forest and go swimming.’
- b. *Řekl jsem, že musíš na nákup a dodělat úkoly.*  
 said AUX.1SG that must.2SG on shopping.ACC and  
*dodělat úkoly.*  
 PF.finish homework.ACC  
 ‘I said that you have to go shopping and finish your homework.’

We thus conclude that the conjoined structures in (17) do not necessarily imply the presence of two VPs, and therefore that of a null verb GO in the conjunct containing the directional PP.

### 3.4. Covert modality

Finally, Marušič and Žaucer (2005) show for Slovenian that infinitival *wh*-clauses, that get some sort of modal interpretation (cf. Bhaat 2000), can also occur with no overt verb and a directional or purpose PP. Assuming that a clause should not exist without a verb, Marušič and Žaucer (2005) claim again that we have to postulate a null verb GO to explain such examples:



- (18) Slovenian
- a. *Tinčku so pokazali kako do štačjona.*  
 Tinček.DAT AUX.3PL.PST showed.3PLhow to station  
 'They showed Tinček how to go to the train station.'
- b. *Se zdaj ne ve kako z biciklom po vino.*  
 still now not knows how with bike for wine  
 'He still doesn't know how to go and get wine by bike.'

In Czech, both directional and purpose PPs like *na nádraží* ('to station') and *pro chleba* ('for bread') respectively may also appear in embedded interrogative clauses without an overt motion verb:

- (19) Czech
- a. *Zptej se ho, kudy na nádraží.*  
 ask REFL he.DAT which.way to station  
 'Ask him how to get to the station.'
- b. *Nevím, jak v tom sněhu pro chleba.*  
 NEG.know.1SG how in this snow for bread  
 'I don't know how to go and get some bread in this snow.'

Moreover, these PPs may also appear in independent non-sentential interrogative fragments like in (20). In contrast, these contexts do not license static PPs that require the presence of the copula *být* ('be') both in embedded wh-clause in (21a) and in non-sentential interrogative fragments in (21b).

- (20) Czech
- a. *Kudy na nádraží?*  
 which.way to station  
 'Which way should we take to go to the station?'
- b. *Kam pro chleba?*  
 where for bread  
 'Where should we go to get some bread?'

- (21) Czech
- a. *Zeptej se ho, kde je nádraží. / \*kde*  
 ask REFL he.DAT where is station where  
*nádraží.*  
 station  
 ‘Ask him where the station is.’
- b. \* *Kde nádraží? / Kde je nádraží?*  
 where station where is station  
 ‘Where is the station?’

In contrast, directional PPs are not able to function as predicates in contexts without modality (and without tense and agreement, see section 4) even if we could postulate a null GO:<sup>10</sup>

- (22) *Já jdu na nádraží. / \*Já na nádraží.*  
 I go.1SG to station I to station  
 ‘I am going to the station.’

Assuming thus that wh-contexts contain some covert modality, we can suppose that this is precisely this covert modality that licenses directional and by extension purpose PP (but not the static PPs), and not a motion verb. We thus conclude that these contexts do not necessarily require the presence of a null verb GO.

### 3.5. Summary

In this section, we presented the main arguments Marušič and Žaucer (2005) give in favor of a null verb GO in structures with directional adverbials. However, we do not think them very convincing at least for Czech. The impossibility to have two contradictory temporal adverbials implies that the structures under discussion involve only one event, which does not require the presence of a null V of movement. The fact that

---

<sup>10</sup> Cf. MacShane (2000) who shows that this is also true for Polish, but not for Russian.

purpose PPs behave like directional PPs is not surprising because the purpose PPs imply a place where we have to go in order to get something. The data only show that directional and purpose PPs may occur in verbless contexts provided these contexts are in some way modal. Finally, the coordination of a PP and a VP after the modal verbs does not imply VP-conjunction either, because coordination does not require strict categorial identity. VP and directional PP can be conjoined because they are semantically predicative phrases. In the next section, we propose another line of reasoning that allows for a GO-less analysis of the structures we deal with.

#### 4. For a GO-less analysis

We have seen above that directional and purpose PPs may appear in *wh*-contexts, which contain some sort of covert modality. The same observation can be made for exclamative contexts, which are also associated with modality (cf. Le Querler 1996). As shown in (23), exclamatives allow for directional but not for static PPs:

- (23) a. *Domů!* / *Do postele!*  
           home.DIR in bed.GEN  
           ‘Go home!’/ ‘Go into bed!’  
       b. \**Doma!* / \* *V posteli!*  
           home.LOC in bed.LOC

We will henceforth assume that PPs with motion and purpose<sup>11</sup> interpretation can be licensed by modality. The question arises how these PPs can be licensed in declarative clauses which are not by themselves modal and require an overt verbal element to bear agreement and tense features<sup>12</sup>. This is the reason why we have to use the copula *být* ‘be’ with nominal,

<sup>11</sup> Within a game e.g., shouting *Na ně!* ‘for them’ means ‘Let’s run on the enemy!’.

<sup>12</sup> Syntactically declarative clauses can be used as assertions or as yes-no questions.

adjectival or adverbial predicates; the copula ‘be’ however licenses only adverbials with non-motion interpretation, *i.e.* static PPs or source PPs like in (24).

(24) Czech

*Jsem v lese. / Jsem z vesnice. / \*Jsem do vesnice.*  
 am in forest.GEN am from village.GEN am to village.GEN  
 ‘I am in the forest.’ / ‘I am from a village.’ / ‘\*I am to a village.’

We claim that the PPs under discussion can be inserted into the syntactic structure in two ways. They can first be selected by lexical verbs of movement which do not by themselves express the direction nor the goal of the movement. The motion verbs in Germanic and Slavic languages actually express the manner but not the direction (cf. Talmy 1991), see the verbs in (25). The directional or goal PPs function thus as part of a complex predicate whose meaning is ‘to move in some manner  $x$  to some place  $y$  in order to get  $z$ ’.

(25) Czech

*jít / jet / běžet / letět / hopkat do lesa na maliny.*  
 walk ride run fly skup to forest for strawberries

The predicate expressing a motion event contains thus three variables:  $x$ ,  $y$  and  $z$ . The variables  $y$  and  $z$  can be easily left unexpressed because leaving out the place or the goal variable allows still to obtain a clause with a verbal predicate. In contrast, if we leave out the manner variable, which is expressed on the lexical verb itself, we end up with a verbless predicate, and the sentence will be ruled out by the grammar. We propose however that we may insert these PPs into syntax without a verb of movement, *i.e.* as predicates, provided that they are supported by some verbal element. Since directional/goal PPs can be semantically licensed by modality, the verbal element required as a support for these PPs would be a modal verb. The modal verb would thus function as a verbal support whose role is to establish the predicative relation between the subject

and the non-verbal predicate, to bear tense and agreement features, and to bear negative prefix *ne-* expressing sentential negation. This proposal implies that modal verbs always combine with a predicative phrase, which could be verbal or non-verbal, the latter one being limited to directional and goal PPs, see (26).<sup>13</sup> The verbal predicate would not be limited to the verbs of movement.

- (26) a. [TP ... [AgrP ... [NegP [ModP ... [PredP ...]]]]]  
 b. [TP *nemusím* [AgrP *ne-t<sub>mod</sub>* [NegP *ne-t<sub>mod</sub>* [ModP *t<sub>mod</sub>* [PredP [PP *pro Adama*]]]]]]]

The possibility for a modal verb to combine either with a verbal predicate denoting a movement event or with a directional/goal PP predicate implying a movement event leads to two different surface structures that seem semantically equivalent, see (27). However, the manner variable, which is explicitly expressed on the verb of movement, is missing in structures with directional/goal PP predicates; it can only be inferred from the subject or the situation to which the adverbial predicate applies.

- (27) a. *Už jsou čtyři, musím (do)jít/ (do)jet/ běžet pro Adama.*  
 already are four must.1SG PF.go PF.drive run for Adam.ACC  
 'It's already 4 o'clock, I have to go and pick Adam (from kindergarden).'
- b. *Už jsou čtyři, musím pro Adama.*  
 already are four must.1SG for Adam.ACC  
 'It's already 4 o'clock, I have to go and pick Adam (from kindergarden).'

Thus, in (27a), the structure containing an overt verb of movement denotes a movement event in which a human subject has to go in some manner to the place where Adam actual-

<sup>13</sup> See Hansen (2000) and Gruet-Skrabalova (2019) for other types of non-verbal complements of modal verbs.

ly in order to achieve the goal of picking him up from that place. The manner is explicitly given by the verb: to walk, to drive, to run. In (27b), the structure containing only the goal PP implies that there is a movement event in which a human subject has as goal to pick up Adam from the place where Adam actually is. The hearer can however infer the manner of achieving this goal from the situation or from his informational background (e.g. usual situation).

## 5. Conclusion

In this paper, we have argued that structures where strict modal verbs combine at surface with an adverbial complement do not require to postulate the presence of a null verb GO. We have shown that these structures refer to a single semantic event, denoted by a single predicate, which can contain verbal or non-verbal material. We have also shown that covert modality licenses directional and goal PPs in *wh*-clauses, non-sentential *wh*-fragments and exclamative clauses. We have therefore argued that modality may license such PPs also in declarative sentences provided that there is an overt verbal element able to bear functional features. In our proposal, we have put forward that directional and goal PPs refer to place and goal variables that are parts of a movement event. We have proposed that these PPs can be inserted into the syntactic structure either as complements of a lexical verb of movement, or as non-verbal predicates. The latter insertion requires that two conditions be met within the declarative clause: presence of modality and presence of a verbal head. These two conditions are successfully met in sentences with modal verbs. Modal verbs are functional heads that express semantic modality and that are able to bear tense and agreement features, and also to support the negative prefix *ne-* expressing sentential negation. The adverbial predicate in these structures implies a movement event, whose manner variable can be inferred from the subject, the situation or the hearer's knowledge.

## References

- Barbiers, Sjef (1995). *The Syntax of Interpretation*. The Hague: Holland Academic Graphics.
- Bhatt, Rajesh (2000). *Covert Modality in Non-finite Contexts*. PhD dissertation, University of Pennsylvania.
- Bayer, Samuel (1996). "The coordination of unlike categories". *Language* 72: 579–616.
- Gruet-Skrabalova, Hana (2019). "Complémentation des verbes modaux en tchègue". *Travaux linguistiques du CerLiCo* 30: 63–86.
- Hansen, Björn (2000). "The German modal verb *müssen* and the Slavic languages: The reconstruction of a success story". *Scando-Slavica* 46: 77–93.
- Le Querler, Nicole (1996). *Typologie des Modalités*. Caën: Presses Universitaires de Caën.
- McShane, Marjorie J. (2000). "Verbal ellipsis in Russian, Polish and Czech". *Slavic and East European Journal* 44: 195–233.
- Marušič, Franc, Rok Žaucer (2005). "On phonologically null verbs GO and beyond". In: Sylvia Blaho, Luis Vicente, Erik Schoorlemmer (eds.). *Proceedings of Console XIII*. Leiden: University of Leiden, 231–248.
- Marušič, Franc, Rok Žaucer (2006). "On the intensional *feel-like* construction in Slovenian". *Natural Language and Linguistic Theory* 24: 1093–1159.
- van Riemsdijk, Henk (2002). "The unbearable lightness of GOing". *Journal of Comparative Germanic Linguistics* 5: 143–196.
- Ross, John R. (1976). "To have *have* and not to have *have*". In: Mohammad A. Jazayery, Edgar C. Polome, Werner Winter (eds.). *Linguistic and Literary Studies in Honor of Archibald A. Hill*. Lisse: Peter de Ridder Press, 265–270.
- Talmy, Leonard (1991). "Path to realization: A typology of event conflation". In: Laurel A. Sutton, Christopher Johnson, Ruth Shields (eds.). *Berkeley Working Papers in Linguistics*. Berkeley, CA: BLS Press, 480–519.
- Travis, Lisa (2000). "Event structure in syntax". In: Carol Tenny, James Pustejovsky (eds.). *Events as Grammatical Objects*. Stanford, CA: CSLI Publications, 145–185.

Hana Gruet-Skrabalova  
Université Clermont Auvergne  
UFR Lettres, Cultures  
et Sciences Humaines  
29 bd Gergovia  
63001 Clermont-Ferrand  
France  
[hana.gruet-skrabalova@uca.fr](mailto:hana.gruet-skrabalova@uca.fr)



## **On *wh*-movement in echo-questions and crosslinguistic variation**

EKATERINA CHERNOVA

*Received 05.09.2021,  
received in revised form 25.11.2021,  
accepted 26.11.2021.*

### **Abstract**

This paper examines echo *wh*-questions, a rather understudied phenomenon even in extensively described languages such as English. In particular, it focuses on a particular type of echo questions, such as those made in response to a previous declarative (e.g., *–Mary said {mumble}./ –Mary said what?*) or a previous *wh*-question (e.g., *–Who said {mumble}?!/ –Who said what?*). Such structures are examined from a comparative perspective, analyzing data from three different languages regarding Multiple *wh*-Fronting: English *vs.* Russian, with attention to Spanish. On the one hand, this paper considers the key, cross-linguistically common features of echo questions and discusses their underlying derivational structure. On the other hand, contrary to the standard assumptions that echo questions necessarily require *wh*-in-situ, this paper focuses on the availability of different options of overt echo *wh*-movement among the languages under consideration. It is argued that in echo questions, similarly to what happens in canonical interrogatives, *wh*-movement proceeds successive-cyclically and is subject to parametric variation.

**Keywords**

echo questions, *wh*-movement, multiple *wh*-fronting, phases, parametric variation

**O ruchu *wh* w pytaniach echo i różnicach językowych****Abstrakt**

Ten artykuł analizuje pytania echo z elementem *wh*, nie do końca zbadanego zjawiska nawet w obszernie opisanych językach, takich jak angielski. W szczególności koncentruje się na konkretnym typie pytań echo, a mianowicie takich, które zadano w odpowiedzi na poprzednie zdanie oznajmujące (np. –Mary said {mumble}./ –Mary said co?) lub na poprzednie pytanie *wh* (np. – Who said {mumble}?/ –Who said what?). Struktury takie są badane z perspektywy porównawczej, w zakresie *Multiple wh-Fronting*. Analizowane są trzy języki: angielski vs. rosyjski, z uwzględnieniem hiszpańskiego. Z jednej strony niniejszy artykuł rozważa kluczowe, wspólne dla wielu języków, cechy pytań echo i omawia ich podstawową strukturę derywacyjną. Z drugiej strony, w przeciwieństwie do standardowych założeń, że pytania echo koniecznie wymagają *wh*-in-situ, niniejszy artykuł koncentruje się na dostępności różnych opcji ruchu *wh* w strukturze powierzchniowej wśród rozważanych języków. Twierdzi się, że w pytaniach echa, podobnie jak w pytaniach kanonicznych, ruch *wh* przebiega sukcesywnie cyklicznie i podlega zmienności parametrycznej.

**Słowa kluczowe**

pytania echo, ruch *wh*, *multiple wh-fronting*, różnica parametryczna, fazy

**1. Introduction**

This paper aims to shed some light on the syntactic behavior of echo *wh*-questions (henceforth *wh*-EQs), which are interrogative sentences produced as an immediate response to

a previous utterance requesting information about some portion of the stimulus that has been missed or unheard. This type of question is exemplified below, in (1b) (hereafter, the echo-introduced *wh*-phrases appear in small caps):

- (1) a. *Mary had tea with Dracula.*  
 b. *Mary had tea with WHO?* [from Sobin 2010:132]

As stated in Sobin (2010:132), “EQs present considerable challenges to theories of interrogative syntax predicated on the behavior of non-EQ interrogatives”, as they behave in a quite unusual way. For instance, EQs appear to counterexemplify some general statements about the formation of true, non-EQ questions such as the obligatoriness of *wh*-movement or the sensitivity to Superiority effects, which are illustrated in (2) for English:

- (2) a. *What<sub>i</sub> did John say t<sub>i</sub>?*  
 b. *Who<sub>1</sub> said what<sub>2</sub>?*  
 c.\**What<sub>2</sub> did who<sub>1</sub> say?*

However, English *wh*-EQs are immune to the obligatory *wh*-movement and the consequent verb raising, as we have seen in (1b) (*cf.* (2a)). Moreover, when an EQ has more than one *wh*-word, Superiority effects can be easily violated without resulting in ungrammaticality (see (3b); *cf.* (2c)):

- (3) a. *What did Dracula drink at Mary's party?*  
 b. *What did WHO drink at Mary's party?* [from Sobin 2010:132]

For these and other reasons, EQs tend to be seen as a non-syntactic phenomenon, as they systematically disobey the general rules of question formation. It seems “unprofitable to attempt to integrate them into the analysis of the more usual types of questions” (Culicover 1976:73), because “the grammatical rules of the language should not generate them”

(Cooper 1983:149). However, here I agree with Sobin (2010: 131), who made the first attempt to capture English EQs in terms of generative syntax, arguing that EQs, being “in the realm of ‘automatic’ and ‘untutored’ knowledge”, are “of great interest and relevance to analyses of question formation”.

Different from true, canonical questions, EQs do not request for new information: “instead they are requests for confirmation of something someone has heard” (Carnie 2006:340) or has not understood. EQs are strongly bound to the previous discourse and, thus, sometimes are referred to as *backward citations* (Escandell 1999) or *reprise questions* (Ginsburg and Sag 2000). To illustrate this point, observe again the previous example, (3). The questioner in (3b) cannot hear a part of the *wh*-question in (3a) pronounced by their interlocutor (suppose, they cannot hear *Dracula*). So, the speaker formulates a *wh*-EQ, in which the echo *wh*-word *who* substitutes the unheard portion of the utterance in (3a), the rest of the utterance being reproduced without changes, including the *wh*-word *what*.

So, echo *wh*-words are referential items, in the sense that they ask about a referent which has been already mentioned in the immediately previous discourse. By using as a question an *undeformed* utterance (i.e., a question with *wh*-in-situ), the speaker presents themselves “as being unable to complete the utterance in a satisfactory way” (Fiengo 2007:76) and asks the addressee to repeat a missing bit of language and to assign a value to the echo *wh*-word.<sup>1</sup>

---

<sup>1</sup> Of course, EQs can be produced in immediate response to an utterance not only in order to request for repetition, as in (2b), but also to express speaker’s surprise, as in (i) (following Bartels (1997), I call the former type *unheard* EQs and the latter, *amazement* EQs):

- (i) a. A: *We’re going to Pakistan on vacation.* (English)  
 b. B: *You’re going WHERE on vacation?!*  
 c. A: *Well, the nature is beautiful there.* [from Šimík 2009: 5]

In this paper, I restrict my attention only to unheard EQs, which can be considered interrogative constructions both from syntactic and semantic points of view (i.e., they seek to reduce the speaker’s ignorance about some missed portion of the stimulus, denoted by a *wh*-word, under which the

In this paper, I mainly restrict my attention to the *wh*-EQs containing two *wh*-items, like those in (3b), in which one *wh*-word is “inherited” from the utterance and the second one is echo-introduced. Moreover, I put forth some novel evidence supporting a view that EQs are syntactic phenomena, underlined by a particular syntactic structure: namely, one involving two CP-levels (see Sobin 2010; Chernova 2015, 2017). As it will be argued, such a view allows accounting for several striking properties of EQs without appealing to purely discursive notions.

Here I focus on some key properties of EQs across two languages with different *wh*-fronting strategies: English and Russian. Additionally, I bring into discussion Spanish *wh*-EQs, in order to support some points of my argumentation. Consider the examples below. As is well-known, Russian, (4c), differs from English and Spanish, (4a) and (4b) respectively, in that the former exhibits obligatory Multiple *wh*-Fronting (hereafter, MWF) in questions with more than one *wh*-word. In other words, all *wh*-items must undergo movement in Russian (as generally in Slavic; see Rudin 1988; Richards 2001; Bošković 2002; among many others):

- (4) a. Who<sub>1</sub> sees whom<sub>2</sub>?  
 b. Spanish  
 ¿Quién<sub>1</sub> ve a quién<sub>2</sub>?  
 who.NOM sees who.ACC  
 ‘Who sees whom?’  
 c. Russian  
 Kto<sub>1</sub> kogo<sub>2</sub> vidit?  
 who.NOM who.ACC sees  
 ‘Who sees whom?’

---

proposition contained within the utterance is true). In contrast, the meaning of an amazement EQ is rather similar to an exclamative: in (ib), the speaker B knows exactly what has been said, however, in their opinion, Pakistan is the least expected place to go on vacation.

Recall that the few existent studies on EQs have argued that this type of interrogatives does not exhibit overt *wh*-movement (see Noh 1998; Iwata 2003; den Dikken 2003; Fiengo 2007; Sobin 2010; among others). Consider the English example in (5), where the questioner cannot hear a part of the previous utterance, a *wh*-question, (5a), and asks for clarification. Observe that in the EQ the echo *wh*-word *who* can only appear in-situ (notice that the apparent Superiority violation does not lead to ungrammaticality in this example), (5b) *vs.* (5c).<sup>2</sup> (In offering examples, I will signal an utterance with *U* and an echo response to it with *EQ*; the unheard portion of the discourse is signaled with {mumble}.)

- (5) a. U:    What did {mumble} buy yesterday?  
       b. EQ:  What did WHO buy yesterday?  
       c. EQ: \* WHO bought what yesterday?

However, as discussed in Chernova (2015), Russian EQs do allow overt *wh*-movement in EQs. Compare the following examples from Russian with what we saw for English in (5):

- (6) Russian  
     a. U:    Kogo    *udaril* {mumble}?  
             who.ACC hit  
             ‘Whom did {mumble} hit?’  
     b. EQ:  Kogo    *udaril* *KTO*?  
             who.ACC hit    who.NOM  
             ‘Whom did hit WHO?’  
     c. EQ:  Kogo    *KTO*    *udaril*?  
             who.ACC who.NOM hit

---

<sup>2</sup> A reviewer suggests an interesting example of partial *wh*-movement in English EQs, which sounds marginal (although not completely odd) in contexts where there is a need to recover some missed part of the previous stimulus:

- (i) a. U:    Where did they buy {mumble} after the meeting?  
       b. EQ:?? Where WHAT did they buy after the meeting?

I leave a detailed account of such cases of partial *wh*-movement in English questions as well as its comparison with Russian and/or Spanish for future research.

- d. EQ: ? *kto* kogo *udari*?  
 who.NOM who.ACC hit

As shown in (6), in Russian, in addition to the *wh*-in-situ option, (6b), the echo *wh*-phrase (here, *kto* ‘who.NOM’) has two other possible landing sites. On the one hand, it can undergo *partial wh*-movement (Fanselow 2005)<sup>3</sup> to some preverbal position, below the *wh*-word “inherited” from the utterance (here, *kogo* ‘who.ACC’), (6c). On the other hand, the echo *wh*-word can also appear at the leftmost position, above the utterance’s *wh*-word, (6d).<sup>4</sup>

Finally, apart from English and Russian EQs, I also consider here some data from Spanish, a language with restricted availability for multiple *wh*-movement under certain licensing contexts (see Etxepare and Uribe-Etxebarria 2005; Uriagereka 2005; Gallego 2017). Observe from the examples below that in this language there are at least two available options for the echo-inserted *wh*-word:<sup>5</sup>

<sup>3</sup> Following Fanselow’s (2005:439) terminology, I assume that “movement is *partial* whenever the phrase has been displaced but its final landing site is below the relevant position”.

<sup>4</sup> As reported in Chernova (2015), for some Russian speakers the in-situ position of the echo *wh*-word, as in (5b), is dispreferred over any other option with movement. There is also some variation regarding the leftmost position of the echo-introduced *wh*-phrase, (5d): many speakers judge it as marginal although possible, while others consider it perfectly acceptable; finally, few speakers reject it.

<sup>5</sup> Notice that the EQs in (6) are different from another type of Spanish interrogatives, as in (i), which have received attention in the studies of Spanish questions with *wh*-in-situ (see Jiménez 1997; Uribe-Etxebarria 2002; Etxepare and Uribe-Etxebarria 2012; Reglero 2007):

- (i) a. A: *Mi padre, mi madre y yo fuimos a la tienda a comprar huevos, leche y café. Mi madre compró los huevos.* (Spanish)  
 ‘My father, my mother and I went to the store to buy eggs, milk and coffee. My mother bought the eggs.’  
 b. B: *Y tu padre compró ¿qué?*  
 ‘And your father bought what?’ [from Etxepare and Uribe-Etxebarria 2005:10]

Equally to EQs, such questions are necessarily linked to the previous discourse, but, unlike EQs, they do not ask about what has been said; rather, they ask about a strong presupposition following from the context.

- (7) a. U: Qué ha leído {mumble}?  
 who.ACC has read  
 ‘What has {mumble} read?’
- b. EQ: Que qué ha leído QUIÉN?  
 that what.ACC has read who.NOM  
 ‘What has WHO read?’
- c. EQ: ??Que qué QUIÉN ha leído?  
 that what.ACC who.NOM has read
- d. EQ: \*Que QUIÉN qué ha leído?  
 that who.NOM what.ACC has read

Similar to what we have seen for English, Spanish speakers show a strong preference for the *wh*-in-situ option, (7b), and unanimously judge as ungrammatical the possibility of overt echo *wh*-fronting into the leftmost position, (7d). However, differently from English and similarly to Russian, in Spanish, the echo *wh*-item can undergo partial movement into some immediately preverbal position, below the *wh*-word “inherited” from the previous utterance, (7c).<sup>6</sup>

Let us summarize the data seen so far. On the one hand, all three languages under consideration allow the in-situ option in *wh*-EQs (although, it is dispreferred in Russian). On the other hand, partial *wh*-movement to some immediately preverbal position is allowed both in Spanish and Russian, but it is blocked in English. Finally, overt echo *wh*-movement to the left edge of the interrogative clause is acceptable only in Russian. This is summarized below in Figure 1.

---

<sup>6</sup> Although both movement options are usually judged as odd by Spanish speakers (with different degrees of marginality), many of my informants notice an interesting contrast. Namely, questions with partial movement of the echo *wh*-word, (6c), sound certainly better than the one with complete *wh*-movement, (6d). For a detailed discussion of Spanish *wh*-EQs, see Chernova (2017).



	English	Spanish	Russian
$Wh_{EQ}$ -in-situ	✓	✓	✓
Partial movement ( $wh_U > wh_{EQ}$ )	✗	✓	✓
Overt movement ( $wh_{EQ} > wh_U$ )	✗	✗	✓

**Figure 1**Echo *wh*-movement across languages

In this paper, I mainly discuss why languages like Russian and Spanish do allow *wh*-movement in EQs and other languages like English do not. This paper aims to propose a syntactic account that can capture the attested parametric differences regarding movement.<sup>7</sup> As already mentioned, I focus here on EQs with two *wh*-words: one comes from the stimulus, and the other one is echo-introduced. I propose that similar to any type of syntactic movement, echo *wh*-movement proceeds successive-cyclically. However, its legitimacy is restricted by certain well-established parametric differences among *wh*-fronting languages. Namely, I argue that it depends on two main factors: (i) the clause-typing properties of the echoed utterance (declarative, interrogative, etc.) and (ii) the number of escape hatches out of phases (such as CP) available in a particular language.<sup>8</sup> I show that we can get a deeper understanding of even such a striking (and apparently ‘non-syntactic’) phenomenon such as EQs if we analyze it comparatively (here, between three typologically different *wh*-fronting languages) and under a unifying theory: namely, Cable’s (2010) Q-based approach to the interrogative syntax.

<sup>7</sup> Due to space restrictions, in this paper I will not address in detail the *wh*-in-situ option, available for all three languages. For a detailed discussion of how EQs with *wh*-in-situ can be accounted for in light of Q-based theory, the reader is referred to Chernova (2015).

<sup>8</sup> As is well known, apart from the standard phase heads C and *v* (Chomsky 2000, 2001), in some languages additional phase domains may be activated: e.g., TP in Romance (see Gallego 2010) or AspP in Slavic (see Dya-konova 2009; Chernova 2015). That is, I claim that echo *wh*-movement crucially depends on whether a language has available escape hatches for extraction of the *wh*-word from the lower domains into the highest level, CP<sub>EQ</sub>. I will turn back to this idea in section 4.

The paper is organized as follows. In section 2, I review several key properties of *wh*-EQs, which distinguish them from canonical *wh*-questions, and propose that EQs have a particular syntactic structure. In section 3, I discuss three theoretical assumptions I make to account for the echo-puzzle: namely, Sobin's (2010) proposal on the double-CP structure underlying this type of questions; then Cable's (2010) Q-based approach to the derivation of canonical *wh*-questions; and, finally, the idea that languages can differ concerning what portion of the structure becomes a phase domain (see Gallego 2007, 2010; den Dikken 2007, among others). Afterward, in section 4, I offer an account that allows us to capture the attested variation regarding echo *wh*-movement in a uniform manner. First, I address EQs with full echo *wh*-extraction, to the leftmost position of the clause (available only in Russian, (6)), and then I consider EQs resorting to partial *wh*-movement (allowed in Russian and Spanish, (6)-(7)). Section 5 concludes this paper.

## 2. Key syntactic properties of *wh*-EQs

Many of the challenging properties of EQs have been previously reported in the literature (e.g., see Sobin 1978, 1990, 2010; Parker and Pickeral 1985; Dumitrescu 1992; Noh 1998; Ginsburg and Sag 2000; Escandell 1999, 2002; Iwata 2003; Fiengo 2007; Sudo 2007; Vlachos 2012; Chernova 2013, 2015, 2017; among others). Here I consider only a few of them, the most relevant ones for the topic at hand.<sup>9</sup>

As already mentioned in the previous section, perhaps, one of the most well-known and, at the same time, striking features of EQs is that they preserve the clause-typing properties of the sentence they "echo" (see Sobin 2010; Noh 1998; Escandell 2002; among others). As we have already seen for English in (5) (consider also (8) below), when an EQ repeats a previous *wh*-question, it has to maintain the *wh*-interro-

---

<sup>9</sup> For a detailed review of the echo-features, the interested reader is referred to Sobin (2010) and Chernova (2015).

gative morphosyntactic features of the echoed utterance (here, the fronted *what* and the raised auxiliary *did*). In addition, the EQ introduces its own syntactic features (here, the echo *wh*-word *who*). Although the resulting structure clearly violates Superiority, the structures in (5b) and (8b) are grammatical contrary to their counterparts in (c).

- (8) a. U: What *did* *{mumble}* drink at Mary's party?  
 b. EQ: What *did WHO* drink at Mary's party?  
 c. EQ: \* *WHO<sub>i</sub> t<sub>i</sub> drank what at Mary's party?* [from Sobin 2010:132]

Similarly, the EQ in (9b), based on a previous polar question, has to preserve the yes/no nature of the echoed utterance, (9a); so, the echo *wh*-word must remain in-situ:

- (9) a. U: *Did Mary have tea with {mumble}?*  
 b. EQ: *Did Mary have tea with WHO?*  
 c. EQ: \* *WHO<sub>i</sub> did Mary have tea with t<sub>i</sub>?* [from Sobin 2010:132]

However, observe that in the case of “echoed” declarative sentences, as in (10), the echo inserted *wh*-word can either appear in-situ, as in (10b), or undergo overt *wh*-movement into the left periphery of the question (with a consequent raising of the auxiliary *did*), as shown in (10c):<sup>10</sup>

---

<sup>10</sup> The fact that a declarative sentence in (9a) can be echoed both with *wh*-in-situ, (9b), and *wh*-ex-situ, (9c), leads Sobin (2010:132) to conclude that the latter is not a syntactic EQ, but rather an instance of what he calls *pseudo EQs*, “simply normally formed questions but with EQ intonation (a strong upward intonational contour)”. According to Sobin, this type of sequences is only possible in response to a declarative utterance, as, according to him, EQs must preserve the clause-typing features of the sentence they echo. However, here I analyze cases like (9c) from a different angle: In languages like English, only declarative utterances can give rise to EQs with overt movement; while in languages like Russian, the option of the explicit echo *wh*-movement is not restricted only to declaratives (see also Chernova 2015, 2017). I will come back to this issue later in this paper.

- (10) a. U: *Mary had tea with {mumble}?*  
 b. EQ: *Mary had tea with WHO?*  
 c. EQ: \* *WHO<sub>i</sub> did Mary have tea with t<sub>i</sub>?* [from Sobin 2010:132]

Generally, an EQ can reproduce any kind of utterance: for example, an exclamative, (11), or an imperative, (12):

- (11) a. U: *What a great pleasure this is!*  
 b. EQ: *What a great WHAT this is?*  
 (12) a. U: *Go to see the archaeologist.*  
 b. EQ: *Go to see WHAT/WHO?*

As the reader may observe, the strategy of “echoing” is broadly always the same: an EQ repeats the stimulus and replaces the unheard portion by a *wh*-word.<sup>11</sup> Interestingly, the interrogative clause-typing of the EQ itself co-occurs with the clause-typing of the echoed sentence. In Escandell (2002), this echo-property is called *mood clashes*, while in Sobin (2010) it is named *Comp freezing*. In both cases, the terminology seeks to capture the fact that the resulting EQ conserves the clause-typing markers of the echoed utterance.

Interestingly, an echo-introduced *wh*-word has always the widest possible scope, independently of its position inside the clause (see Sobin 2010; Chernova 2015, 2017). That is, as shown below for English, independently of whether the echo-introduced *wh*-phrase appears in the root clause, (8)-(12), or deeply embedded, (13b), it always receives wide scope and seeks for an answer:

- (13) a. U: *Mary says [that Peter believes [that John is a lover of {mumble}]].*

---

<sup>11</sup> Generally, the natural tendency for an echo *wh*-word is to remain in-situ (at least in English and Spanish), although, as it has been mentioned before, there is certain parametric variation across languages regarding movement in EQs. I will discuss this issue later.

- b. EQ: *Mary says [that Peter believes [that John is a lover of WHO]]?*

Crucially, the root-scope phenomenon is also observed in EQs with two *wh*-items, as in (8b). As is well-known, canonical, non-echo multiple questions presuppose exhaustification of each quantifier, giving rise either to pair-list or single-pair readings (see Higginbotham and May 1981; Hagstrom 1998; Krif-ka 2001). However, neither pair-list nor single-pair interpretations are available in EQs like (7b). For instance, consider (14), where even in the presence of a universal quantifier the EQ only allows the individual reading:

- (14) a. U: *Everybody talked to {mumble}.*  
 b. EQ: *Everybody talked to WHOM?*  
 c. R: *To Mary.*  
 d. R: \* *John talked to Mary (Peter talked to Helen, Bill talked to Nancy...)*

Likewise, in an EQ based on a previous *wh*-question (see example (15) below), only the echo *wh*-word (*who* in (15b)) receives scope, while the *wh*-word inherited from the previous utterance (*what*) requires no response.<sup>12</sup> In fact, the only appropriate answer to (15b) is (15c) (as far as the agent of the action described by the stimulus in (15a) is John indeed):

- (15) a. U: *What is {mumble} going to bring to the party?*  
 b. EQ: *What is WHO going to bring to the party?*  
 c. R: *John*  
 d. R: \* *John is going to bring vodka.*

---

<sup>12</sup> As noted first by Baker (1970), a similar loss-of-scope effect arises in embedded *wh*-questions like (ia), where the embedded *wh*-phrase *what* can receive either narrow scope (in the sense that it does not require any answer), as in (ib), or wide scope, as in (ic) (see also Chomsky 1977a; Pesetsky 1987; Sobin 2010; among others):

- (i) a. *Who knows where Mary bought what?*  
 b. *John does.*  
 c. *John knows where she bought milk, Bill knows where she bought bread...*

- e. R: \* *John is going to bring vodka, Mary is going to bring tequila...*  
 f. R: \* *Vodka.*

Moreover, observe that an echo *wh*-word can appear both inside strong and weak islands (without consequences for the grammatical status of the correspondent sentence). Again, just as in previous examples, it necessarily receives wide scope. This is illustrated below for English, where the island effects are created by sentential subjects, (16), adjuncts, (17), and embedded *wh*-questions, (18), respectively:<sup>13</sup>

- (16) a. U: *Mary left [after John met {mumble}]*  
       b. EQ: *Mary left [after John met WHO]?*  
 (17) a. U: *I think [that to sell {mumble}] would be a mistake.*  
       b. EQ: *You think [that to sell WHAT] would be a mistake?*  
 (18) a. U: *I wonder [who could have {mumble}].*  
       b. EQ: *You wonder [who could have WHAT]?*

Finally, observe that there is an interesting piece of evidence suggesting that EQs do “actively involve syntax” (Sobin 2010:135): the echoed utterance and the correspondent EQ may show different person-agreement features and deictic elements. That is, the content of an EQ is sensitive to the changing discourse roles between the speaker and the addressee (see Dalrymple and Kaplan 2000; Harley and Ritter 2002). This is illustrated below for Spanish:

- (19) a. U: *Me iré a tu casa {mumble}.*  
           CL.1.SGgo.FUT.1SG to your house  
           ‘I will come to your house {mumble}.’

<sup>13</sup> In some sense, EQs in *wh*-fronting languages exhibit similar behaviour to standard, non-echo questions in languages resorting to *wh*-in-situ, (i) (for the latter, see Cheng 1991; Hagstrom 1998; Cheng and Rooryk 2000; Watanabe 2001, 2002; Kishimoto 2005; Cable 2010; among many others):

(i) *Mary-wa* [<sub>DP</sub> [<sub>CP</sub> *John-ni nani-o ageta*] *hito-ni*] *atta-no?*  
 Mary.NOM John-DAT what-ACC gave man-DAT met-Q  
 \*‘What did Mary meet the man who gave \_ to John?’

- b. EQ: ¿(Que) (CUÁNDO) te irás a mi casa  
 that when CL.2.SGgo.FUT.2SG to my house  
 (CUÁNDO)?  
 when  
 'You will leave my house WHEN?'

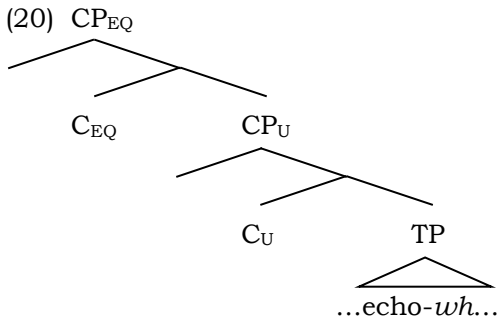
The deictic accommodation (reflected on the verb and the pronouns) corresponds to the two dependents of the participants of the speech act: speaker and addressee (1<sup>st</sup> and 2<sup>nd</sup> person, respectively). These changes are unexpected if we assume that EQs are simply a type of direct quote (e.g., *Mary said: "I am hungry"*). Rather, EQs seem reminiscent of indirect questions (e.g., *Mary said that she was hungry*).

### 3. Derivation of *wh*-EQs: assumptions

#### 3.1. EQs as a double-CP structure

In this paper, I argue that, despite the appearance of being a purely pragmatic phenomenon, unaccountable under any syntactic rule, EQs actually do "actively involve syntax" (Sobin 2010:135). Some of the previously discussed echo-properties, such as co-existence of syntactic features of two different clause-types and wide scope for the echo-inserted *wh*-word independently of its position within a clause, suggest that *wh*-EQs are structurally different from true *wh*-questions.<sup>14</sup> Here I assume a particular echo-structure, originally proposed in Sobin (2010): EQs possess their own, interrogative C head (C<sub>EQ</sub>), in addition to the C head involved in the derivation of the echoed utterance (C<sub>U</sub>). As a result, the syntactic structure of EQs involves two different adjacent CP projections: namely, CP<sub>EQ</sub> asymmetrically c-commands CP<sub>U</sub>. This is schematically represented below:

<sup>14</sup> For a detailed discussion of why EQs should be analyzed as a syntactic phenomenon, see Sobin (2010) and Chernova (2015, 2017).



Under the structure in (20), it is expected that the clause-typing features of the echoed utterance would be preserved in the EQ, through projection of the CP<sub>U</sub> level. However, it does not mean that the derivation of *wh*-EQs somehow implies a “frozen copy” of the utterance’s CP (*contra* Sobin 2010). Rather, I suggest that a C head of the same type as the one of the stimulus is merged during the derivation of *wh*-EQs; consequently, the same type of CP (but, importantly, not the same instance of that CP) is built in the course of the standard, bottom-up derivation. Afterward, an additional functional head is merged into the structure: C<sub>EQ</sub>. It selects CP<sub>U</sub> as a sister and projects a higher, discourse-bound interrogative projection. C<sub>EQ</sub> assigns scope to the anaphoric, echo-introduced *wh*-word within its c-command domain.<sup>15</sup> The higher projection, CP<sub>EQ</sub>, is also responsible for the request-for-repetition meaning of the resulting question. As the derivation proceeds, we obtain a double-CP structure, as in (20).

Following Chernova (2013, 2017), here I assume that Spanish EQs, especially those reproducing a previous yes/no ques-

<sup>15</sup> Bear in mind, however, that Sobin’s proposal has been developed in order to account for English data, with the attested differences between true *wh*-questions (with obligatory *wh*-movement) and *wh*-EQs (always with *wh*-in-situ) (see (7)-(9)). Thus, the widest scope of the in-situ echo *wh*-word is captured through its unselective binding by the highest C<sub>EQ</sub> at a distance, through valuation of the echo-feature (for details, see Sobin 2010: 144-146). However, as already advanced, in this paper, I deal with a different set of data, suggesting that an echo *wh*-item *can* move.



tion, offer an interesting piece of evidence for the double-CP structure proposal. Consider the example below:

- (21) a. U: *¿Ha llegado {mumble}?*  
           has arrived {mumble}  
           ‘Has {mumble} arrived?’  
       b. EQ: *¿(Que) \*(si) ha llegado QUIÉN?*  
           that if has arrived who  
           ‘Has WHO arrived?’

Notice that, in addition to the echo-inserted *wh*-word *quién* ‘who’, the EQ in (21b) exhibits two items that are absent from the original stimulus, the yes/no question, in (21a): *que* ‘that’ (a quotative marker) and *si* ‘whether’ (an interrogative particle). The latter appears only in Spanish EQs based on a previous *yes/no* question: while the lack of *si* would be ungrammatical in (21b), its presence is blocked in EQs built on a previous declarative, (22), or a *wh*-question, (23):

- (22) a. U: *María compró {mumble}.*  
           María bought {mumble}  
           ‘María bought {mumble}.’  
       b. EQ: *¿(Que) (\*si) María compró QUÉ?*  
           that if María bought what  
           ‘María bought WHAT?’  
       (23) a. U: *¿Qué compró {mumble}?*  
           what bought {mumble}  
           ‘What did {mumble} buy?’  
       b. EQ: *¿(Que) (\*si) qué compró QUIÉN?*  
           that if what bought who  
           ‘What did who buy?’

As I argued in Chernova (2013, 2017), *si* is a phonetically realized instance of the interrogative operator Q, which is merged within the CP level of a yes/no question and is responsible for its interrogative interpretation (see Baker 1970). Q tends to be null in Spanish (and also in English) root polar questions, but it becomes phonetically realized in embedded contexts: as *si* in

Spanish (see Rigau 1984; Suñer 1991; Hernanz 2012); *if/whether* in English (Baker 1970) or *se* in Italian (see Rizzi 2001). Thus, in Spanish EQs, as in (21b) (assuming the structure in (20)), *si* is a phonetically realized Q, merged at the specifier of CP<sub>U</sub>; in the root context, the same position is occupied by its phonetically null counterpart.<sup>16</sup> This is schematized below in (24). So, the absence of *si* from the EQs in (22) and (23) is fully expected under such view:

- (24) a. U: [CP<sub>U</sub> Ø C<sub>U</sub> ...].  
 b. EQ: [CP<sub>EQ</sub> [CP<sub>U</sub> *si* C<sub>U</sub> ...]]

Let us now briefly consider the introductory particle *que* ‘that’, which can optionally appear in Spanish EQs, independently of the clause-type of the echoed utterance.<sup>17</sup> I take *que* as a *quotative* marker (see Escandell 1999). In EQs, it signals that the speaker partially reproduces (“quotes”) the words pronounced by her interlocutor in the previous speech turn. The data suggest that this marker is merged within the CP<sub>EQ</sub> level: observe that *que* must always precede *si*:

---

<sup>16</sup> An anonymous reviewer wonders why in English polar EQs *if/whether* (at Spec,QPU under my account) is not present, (i), even though it is in indirect, embedded non-echo polar questions, (ii):

- (i) a. U: Did Mary have tea with {mumble}? (English)  
 b. EQ: (\**If/Whether*) did Mary have tea with WHO?  
 (ii) John asked *if/whether* Mary had tea with Dracula.

Perhaps the answer to the aforementioned contrast is that the auxiliary verb undergoes *v-to-T-to-C* movement in English root questions (contrary to their embedded counterparts; see Pesetsky and Torrego 2001, 2004) and acts as a sort of Doubly-Filled Comp Filter (Chomsky and Lasnik 1977). On the other hand, it has been argued that in Spanish root questions, the verb does not rise so high (see Gallego 2007, 2010 and references therein; see also the footnote 22). I leave this issue for future research.

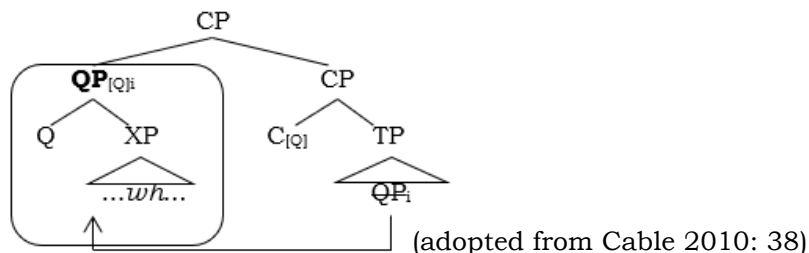
<sup>17</sup> In principle, the particle *que* ‘that’ is optional in Spanish EQs. However, most of my informants note that EQs sound more natural (and are interpreted more easily as *echo*) when *que* is present. This opens an interesting question on which factors affect the degree of optionality of *que*. However, I leave this issue aside for the present.

- (25) a. EQ: ¿Que si *ha llegado QUIÉN?*  
           thatif has arrived who  
           ‘Has arrived WHO?’  
       b. EQ:\*¿Si *ha llegado quién?*  
           if has arrived who  
           ‘Has arrived WHO?’

### 3.2. A Q-particle approach to *wh*-EQs

As for the trigger of *wh*-movement into the left periphery, here I assume the main insights of Cable’s (2010) Q-based theory for true, non-echo *wh*-questions and extend them to *wh*-EQs. According to Cable, movement of a *wh*-word in questions is a secondary effect of Q-movement: in other words, the fronted *wh*-word is not a scope-bearing operator. That is, the syntactic and semantic relations with the interrogative C are established through the help of a Q-particle, which is merged with a *wh*-word (or a larger, *wh*-containing phrase, XP) in its argument position. In languages resorting to *wh*-ex-situ, when Q is merged with a *wh*-phrase, the former takes its sister as a complement and projects its own QP layer, which minimally dominates both items. As a result, the first node endowed with the Q-feature being visible for C is QP. This entails that the attraction of the Q-feature into CP triggers movement of the whole QP (no feature percolation being necessary). This is schematically represented below:

- (26) Derivation of standard *wh*-questions with overt *wh*-movement



Extending Cable's original theory to *wh*-EQs, I argue that the derivation of this type of interrogatives involves three crucial elements: (i) an anaphoric echo *wh*-phrase, corresponding to the unheard portion of the stimulus, merged at the argument position, (ii) a phonetically null discourse-bound interrogative Q-particle ( $Q_{EQ}$ ), merged anywhere in the tree where it c-commands the echo *wh*-word, and (iii) a discourse-related interrogative head  $C_{EQ}$ .<sup>18</sup> Independently of the merging place of the  $Q_{EQ}$ , by the end of the derivation, it must move to the scope position of the question, a syntactic universal. I propose that in *wh*-EQs the scope position for the interrogative operator is in the specifier of  $CP_{EQ}$ . Thus, echo *wh*-movement is triggered by a formal imperfection on the  $Q_{EQ}$ -particle itself and its need to check its feature  $[Q_{EQ}]$  with the head  $C_{EQ}$ .

In principle, under Q-theory, a Q-particle can be merged anywhere in the tree where it c-commands the *wh*-word. Following Cable (2010), I propose that in *wh*-fronting languages the size of a *wh*-containing constituent XP is restricted by the locality-sensitive Agree operation between the Q-morpheme and the echo *wh*-word. Regarding EQs, it means that an echo Q-particle must agree with the echo *wh*-word it c-commands within some local domain.

I suggest that all echo-inserted *wh*-words (recall their anaphoric/referential nature, as opposed to *wh*-words of the ordinary question) enter the derivation bearing a valued instance of the  $[wh]$ -feature, see (27a). Notice, however, that this  $[wh]$ -feature on echo *wh*-elements is different from the standard  $[wh]$  on *wh*-words involved in non-echo questions, as only the former are anaphoric items (hence, they bear a  $[+anaphoric]$

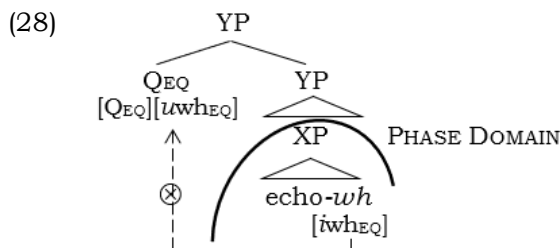
---

<sup>18</sup> In fact, Cable (2010) suggests that different structures might involve different instances of the same category label Q. So, it is natural to assume that a Q-particle involved in the derivation of a true *wh*-question is different from the one involved in the derivation of a *wh*-EQ. One of the main differences between two Qs is that only  $Q_{EQ}$  is anaphoric. Here I signal this property with the index *EQ* on the interrogative Q-particle and the corresponding Q-feature it bears. It also allows us to distinguish between elements involved in the derivation of *wh*-EQs and those found in the derivation of true *wh*-questions.

feature, which I represent here, for simplicity's sake, with an index  $EQ$ ). As for echo Q-particles, assuming Cable (2010), I propose that they must bear a bundle of features, see (27b). In addition to the interrogative feature  $[Q_{EQ}]$ , I suggest that such Q-morpheme also carries an unvalued instance of  $[wh_{EQ}]$ , which forces the Q-particle to agree with the anaphoric *wh*-phrase it c-commands:

- (27) a. Echo *wh*-word:  $\{[iwh_{EQ}]\}$   
 b. Echo Q-particle:  $\{[Q_{EQ}]; [uwh_{EQ}]\}$ <sup>19</sup>

Such Q/*wh*-agreement is subject to locality conditions. In particular, as discussed in Cable (2010), agreement cannot cross into islands and separate spell-out domains. Under such contexts, the unvalued feature on  $Q_{EQ}$  cannot be checked and the derivation fails:

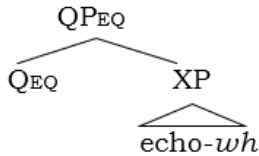


Within Cable's Q-theory, *wh*-ex-situ and *wh*-in-situ structures result from different merging options of Q and its *wh*-containing sister XP. These options are summarized below. On the one hand, the Q-particle can take XP as a complement and project a QP; then the whole complex QP undergoes movement into CP, resulting in *wh*-ex-situ. On the other hand, Q can ad-

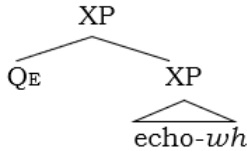
<sup>19</sup> Under this assumption, EQs with *wh*-in-situ (an option allowed in all three languages under consideration) are derived through the help of a particular instance of  $Q_{EQ}$  that does not need to undergo agreement with an echo *wh*-phrase. I do not discuss this option in detail in this paper due to space restrictions. For a detailed account, the interested reader is referred to Chernova (2015).

join to XP; such Q does not project and, as a consequence, it undergoes movement into the scope position of the correspondent question by itself, leaving its sister in-situ. These two options are represented in (29a) and (29b), respectively:

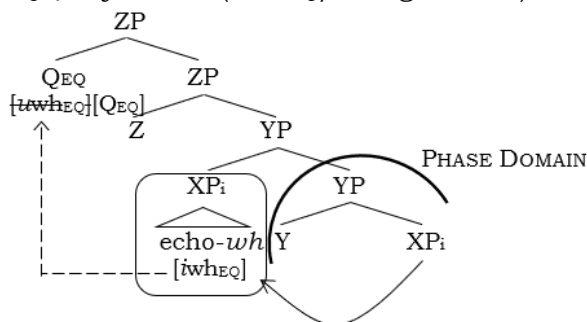
(29) a.  $Q_{EQ}$ -projection (wh-ex-situ)



b.  $Q_{EQ}$ -adjunction (wh-in-situ)



According to Cable, *wh*-fronting languages always resort to Q-projection, while *wh*-in-situ languages like Japanese or Chinese resort to Q-adjunction. However, I propose that even in *wh*-fronting languages, under particular, discourse-bound contexts such as *wh*-EQs, a discourse-bound Q-particle can resort to both merging options. In other words, in *wh*-fronting languages, not all  $Q_{EQ}$ -morphemes need to project. Certain instances of  $Q_{EQ}$  can resort to adjunction, although they still require agreement with the echo *wh*-word within their c-command domain. If the echo Q-particle merges locally, no effect arises at the outcome: after agreement, the echo *wh*-word remains at its base position and  $Q_{EQ}$  undergoes fronting into its scope position on its own. However, if such  $Q_{EQ}$  is initially merged at long distance from its *wh*-containing goal XP, it forces the latter to undergo partial *wh*-fronting from its argument position into the edge of a phase, to become visible for the probe  $Q_{EQ}$ . This type of  $Q_{EQ}$ -adjunction, parameterized for Spanish and Russian (which I will discuss later in this paper), is illustrated below:

(30)  $Q_{EQ}$ -adjunction (after  $Q/wh$ -agreement)

### 3.3. Parametrized points of spell-out

As it is well-established (since Chomsky 2000, 2001), the complement domains of the phase heads, standardly  $\nu$  and  $C$ , become opaque for further operations as a result of being transferred to the external systems (the so-called Phase Impenetrability Condition; PIC).<sup>20</sup> In addition, assuming the idea that  $\nu$ -movement results in the extension of “checking domains” (see Chomsky 1986, 1995), several studies on phases have argued that points of Spell-out are subject to parameterization (see Svenonius 2000; den Dikken 2007; Gallego 2007, 2010; Pesetsky 2007). In other words, languages can differ as to what portion of the structure becomes a phase domain.

The extension of  $\nu P$ 's phasehood in a particular language is parasitic on head movement of  $\nu$  into a higher functional projection since  $\nu$  brings together with it its phasal properties.

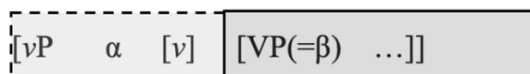
<sup>20</sup> According to Chomsky's (2000, 2001) PIC, once the derivation is done at a given stage, correspondent chunks of structure are spelled-out, thereby becoming inaccessible for the further computation. PIC helps to reduce the computational burden, being a constraint that forces the system to “forget” about transferred portions of the structure. According to Chomsky's (2001:14) version of PIC, the transfer of the complement domain of a phase is delayed until the next phase head is projected; afterwards any further syntactic manipulation of the spelled-out chunk of structure is prohibited:

(i) Given structure [ $ZP Z \dots [HP \alpha [H YP]]$ ], with  $H$  and  $Z$  the heads of phases, the domain of  $H$  [the head of a strong phase] is not accessible to operations at  $ZP$  [the next strong phase]; only  $H$  and its edge [ $\alpha$ ] are accessible to such operations.

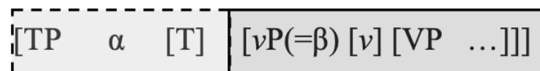




(32) a. English



b. Spanish



Gallego addresses many properties of Spanish (and other Romance languages) which, according to his proposal, are the result of TP being a phase: e.g., *pro*-drop, the fact that subjects can appear both pre- and post-verbally (the formers bearing a topic-like flavor), and the lack of obligatory subject-verb inversion in questions (see (33)), among others.

(33) Spanish (from Gallego 2007: 129)

a. [<sub>CP</sub> *Por qué* C [<sub>TP</sub> *Celia llamó a su hermana*]]?

why Celia called to her sister

'Why did Celia call her sister?'

b. [<sub>CP</sub> *Por qué* C *llamó* [<sub>TP</sub> (*Celia*) *a su hermana* (*Celia*)]]?

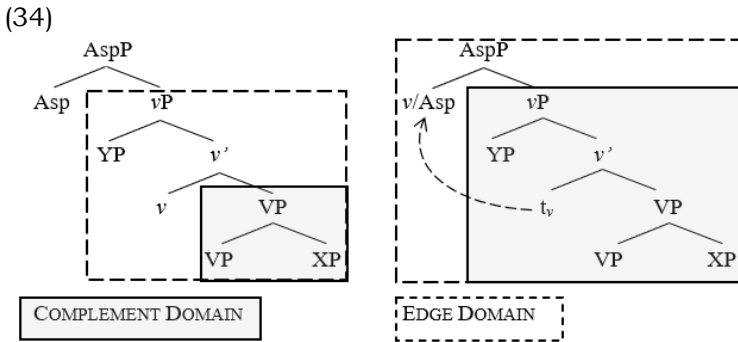
why called Celia to her sister Celia

'Why did Celia call her sister?'

Observe from (33), that, contrary to English questions (with obligatory subject-verb inversion), in their Spanish counterparts the subject (here, *Celia*) can appear both above and below the verb. As discussed in Gallego (2007:129), while the question in (33b) has a standard, out-of-the-blue meaning ('there is a reason *x*, such that Celia did not call her sister because of *x*'), the question in (33a), with a preverbal subject, receives a marked interpretation. Namely, it can mean either 'why was it Celia (and not another person) who called her sister?' or else 'why was it (true) that Celia called her sister?'. According to the author, in (33a) the preverbal subject appears at the edge of TP and, as a consequence, acquires a topic-like flavor. This is a plausible outcome under Chomsky's (2001) claim that discourse-oriented semantics is related to phase

edges. That is, in Spanish the edge of TP can exhibit certain peripheral properties generally attributed to the “standard” phase heads C and  $v$ .<sup>22</sup>

As for Russian, another language with rich morphology, it has been argued that it has an additional phasal projection, AspP (see Dyakonova 2009; Chernova 2015). This idea recasts the well-known fact about the richness of aspectual morphology in Slavic, as opposed to languages like Spanish or English. As is well-known, in Slavic languages, aspectual differences are encoded in verbal morphology, particularly, in a large number of aspectual prefixes (see Svenonius 2004). For instance, Russian has a fairly simple system of tense and a quite complex system of aspect, which means that interpretation of the former is mostly determined by the latter (see Borik 2006; Borik and Reinhart 2004). It has been proposed that in Slavic languages, similarly to Romance, the verb undergoes movement, but “it remains relatively low”, as it “cannot move as high as T” (Svenonius 2004b:6). Namely, the phase head  $v$  moves to Asp (see Svenonius 2004a,b; Ramchand 2004; Bošković 2014). I argue that, as a consequence, the phasal properties of  $v$ P extend to AspP, as represented below (compare with (32)):<sup>23</sup>

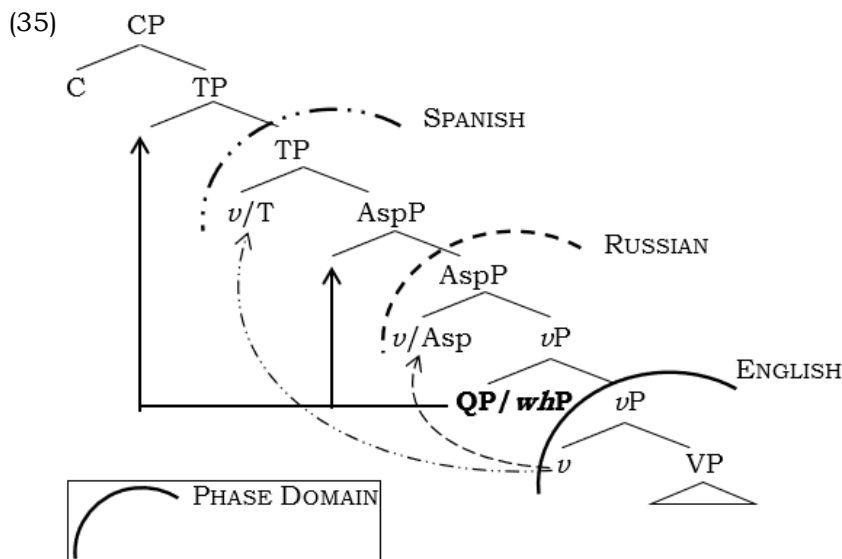


<sup>22</sup> Gallego’s proposal captures the sense of Uriagereka’s (1995) FP, a projection “sandwiched” between CP and IP and encoding discourse-oriented effects.

<sup>23</sup> For a detailed discussion of arguments suggesting  $v$ -to-Asp movement and its application to canonical multiple *wh*-questions the reader is referred to Chernova (2015).

A natural effect of the phase extension is the consequent extension of the phase complement domain, a point that is going to be crucial for our purposes. Recall from (31) that, under PIC, only the head H and the edge of a phase HP are visible for further operations, while the complement of H becomes opaque by being transferred. However, if the phase HP extends to ZP, what used to be the edge of HP turns into the domain of the newly extended phase ZP. As a result, all syntactic objects with unvalued feature(s) are forced to escape the domain of ZP, otherwise, the derivation would crash.

Applying this logic to Russian (and Spanish), I argue that after movement of  $v$  into a higher head, phase extension takes place. That is, the edge of  $v$ P turns into the domain of AspP (in Russian) or TP (in Spanish). Subsequently, all potential goals or elements with any formal imperfection must be removed from the edge of  $v$ P to the edge of the higher, newly constructed phase, in order to be visible for further syntactic operations. In contrast, in English, the phase  $v$ P does not extend and the verb remains low (in  $v$ ). This scenario is schematically represented below for all three languages:



In what follows, I argue that the availability of echo *wh*-movement is parameterized across languages by PIC, as such movement proceeds successive-cyclically out of phase domains through available escape hatches along its path. Partial raising to the preverbal area of echo *wh*-words, an option available in Russian and Spanish EQs, is a consequence of the extension of the  $\nu$ P phasehood and the need of the echo Q-particle to undergo agreement with the echo *wh*-word it c-commands. In EQs, further movement into the highest level, CP<sub>EQ</sub>, proceeds through the edge of the lower CP<sub>U</sub>, which is a phase in all three languages under consideration. As is well-known, the number of available escape hatches out of phases is also parameterized.

#### 4. Accounting for the parametric variation on echo *wh*-movement

As already mentioned through the paper, I claim that EQs in principle allow for both *wh*-in-situ and *wh*-ex-situ strategies, with an intermediate option: partial *wh*-movement (allowed in Russian and, marginally, in Spanish). However, the availability of overt *wh*-movement into the leftmost position of the question is constrained by the clause-type of the echoed utterance. Importantly, there is a crucial observation in Sobin (2010), which I take as a departure point for my argumentation. Consider again (10), repeated below as (36) with some additional items:

- (36) a. U:     *Mary had tea with {mumble}?*  
       b. EQ:    *Mary had tea with WHO?*  
       c. EQ: \* *WHO<sub>i</sub> did Mary have tea with t<sub>i</sub>?*

Sobin observes that English EQs allow for overt *wh*-movement, as in (36c), only when the echoed utterance is declarative, (36a). In effect, recall from our previous discussion that echo *wh*-movement is blocked in other contexts (when the echoed

sentence is either a polar, (37), or a *wh*-question, (38)). In such cases, the only available option for the *wh*-word is to appear in-situ:

- (37) a. U: *Did Mary have tea with {mumble}?*  
 b. EQ: *Did Mary have tea with WHO?*  
 c. EQ:\* *WHO<sub>i</sub> did Mary have tea with t<sub>i</sub>?*
- (38) a. U: *What did {mumble} drink at Mary's party?*  
 b. EQ: *What did WHO drink at Mary's party?*  
 c. EQ:\* *WHO<sub>i</sub> what did t<sub>i</sub> drink at Mary's party?*  
 d. EQ:\* *WHO<sub>i</sub> t<sub>i</sub> drank what at Mary's party?*

Sobin argues that EQs must preserve the syntactic character of the stimulus (under his proposal, by “freezing” the CP of the echoed utterance). Thus, (37c) is ungrammatical because overt *wh*-movement is not compatible with the *yes/no* syntax of the stimulus in (37a). Similarly, movement of the echo *wh*-word *who* in (38c,d) would break the “frozen” CP layer of the echoed *wh*-question in (38a).

However, as we have seen already in (6) (repeated below with additional items as (39)), this prediction does not hold for Russian *wh*-EQs, where the echo *wh*-word can undergo overt movement into the leftmost position even when the CP of the echoed utterance has interrogative syntax. This is exemplified below for EQs built on a previous *wh*-question:

- (39) Russian
- a. U: Kogo *udaril* {mumble}?  
 who.ACC hit  
 ‘Whom did {mumble} hit?’
- b. EQ: Kogo *udaril* KTO?  
 who.ACC hit who.NOM  
 ‘Whom did hit WHO?’
- c. EQ: Kogo KTO *udaril*?  
 who.ACC who.NOM hit
- d. EQ: ? KTO kogo *udaril*?  
 who.NOM who.ACC hit

As shown in (39), in Russian the echo *wh*-word *kto* ‘who.nom’ can appear in-situ, (39b) (just as in English), but it can also undergo movement to an immediately preverbal position, (39c), or even to the left periphery of the question, (39d), above the *wh*-word “repeated” from the stimulus.

We have also seen that in Spanish EQs the option of echo *wh*-movement is neither completely blocked, although it is more restricted than in Russian (see (40) below):

- (40) a. U: Qué ha leído {mumble}?  
 who.ACC has read  
 ‘What has {mumble} read?’  
 b. EQ: Que qué ha leído QUIÉN?  
 that what.ACC has read who.NOM  
 ‘What has WHO read?’  
 c. EQ: ??Que qué QUIÉN ha leído?  
 that what.ACC who.NOM has read  
 d. EQ: \*Que QUIÉN qué ha leído?  
 that who.NOM what.ACC has read

Consider also another example, in (41), where the *wh*-EQs echoes a previous polar question and, in addition, exhibits the quotative marker *que* ‘that’ and the interrogative operator *si* ‘whether’:

- (41) a. U: ¿Has traído {mumble}?  
 Have.2SG brought {mumble}  
 ‘Have you brought {mumble}?’  
 b. EQ: ¿(Que) si he traído QUÉ?  
 that if have.1SG brought what  
 ‘Have I brought WHAT?’  
 c. EQ: ?¿(Que) si QUÉ he traído t<sub>i</sub>?  
 that if what have.1SG brought  
 d. EQ: \*¿(Que) QUÉ si he traído t<sub>i</sub>?  
 that what if have.1SG brought

In Spanish EQs, in addition to the standard *wh*-in-situ option, (40b) and (41b), the echo *wh*-word can also undergo partial

*wh*-movement into some preverbal position, the option (c) in (40) and (41) (similarly to Russian and contrary to English).<sup>24</sup> However, movement into the leftmost position, above *qué* ‘what’ in (40d) and *si* ‘whether’ in (41d) (presumably into CP<sub>EQ</sub>), is blocked (as opposed to Russian).

Thus, our data suggest that the standard assumption on the mandatory *wh*-in-situ for EQs does not hold cross-linguistically. In what follows I offer an account for the overt echo *wh*-movement (available only in Russian) and partial echo *wh*-movement (available in Russian and Spanish) that captures the attested parametric variation uniformly.

#### 4.1. Echo *wh*-movement into the leftmost position

Extending Cable’s (2010) Q-based theory to EQs (see section 3.2), I argue that in EQs with the echo *wh*-word at the leftmost position (e.g., *Who<sub>i</sub> did Mary have tea with t<sub>i</sub>?*, (36c)), what undergoes movement into CP<sub>EQ</sub> is not the echo *wh*-word alone, but rather a complex QP<sub>EQ</sub> projection, which includes the echo Q-morpheme and the *wh*-word.

I claim that echo *wh*-movement proceeds successive-cyclically, through available escape hatches on its way up to CP<sub>EQ</sub>. Following Chomsky’s (2001 *et seq.*) Phase theory, I assume that internal Merge of the fronted *wh*-phrase to the highest CP

---

<sup>24</sup> Recall our observation that EQs with partial fronting of the echo *wh*-word, below the *wh*-item “inherited” from the echoed stimulus, sound quite weird for most consulted Spanish speakers (see the footnote 6). Moreover, an anonymous reviewer brings to my attention an interesting contrast between the following Spanish examples:

- (i) a. ?? ¿Que dónde QUIÉN estaba?
- b. ?(?) ¿Que dónde QUIÉN estaba durmiendo?
- (ii) ¿Qué dónde estaba (durmiendo) QUIÉN?

Although both examples in (i) sound rather odd in comparison with (ii), with *wh*-in-situ, it seems that the question in (ib), with a “heavier” VP, is slightly better than the one in (ia). Although a detailed account of Spanish data falls aside from the scope of this paper, the contrast with their Russian counterpart is noteworthy. Namely, questions with partial movement of the echo *wh*-word have a higher degree of acceptability among Russian speakers than among Spanish speakers. I leave this interesting issue for future research.

node does not proceed in a unique long leap, but rather occurs through the intermediate landing sites, or *escape hatches* (i.e., every specifier along the movement path). Assuming the double-CP structure of *wh*-EQs (schematically represented below in (42)), it is expected that the complete echo *wh*-extraction has to proceed through the edge of CP<sub>U</sub> on its way into CP<sub>EQ</sub>.

$$(42) \left[ \text{CP}_{\text{EQ}} \text{ \_\_\_ } \text{C}_{\text{EQ}} \left[ \text{CP}_{\text{U}} \text{ \_\_\_ } \text{C}_{\text{U}} \left[ \text{TP} \dots \left[ \text{QP}_{\text{EQ}} \text{ } \textit{wh} \right] \right] \right] \right]$$

$\uparrow$  \_\_\_\_\_ |  $\uparrow$  \_\_\_\_\_ |

Under this view, the grammaticality of EQs with full *wh*-extraction crucially depends on the availability of the specifier of CP<sub>U</sub> as an escape hatch out of the phase domain. I argue that this is precisely the reason why the clause-type of the echoed utterance (declarative *vs.* interrogative) plays such an important role for overt echo *wh*-movement.

When an EQ is based on a declarative utterance, the edge of the phase CP<sub>U</sub> is left unfilled; so it can act as an escape hatch for an echo *wh*-word on its way to the edge of the higher CP<sub>EQ</sub>. I claim that this is the reason why the English example in (36c), with overt extraction of the echo *wh*-word (repeated below as (43b)) is grammatical:<sup>25</sup>

- (43) a. U: *Mary had tea with {mumble}*.  
 b. EQ: [CP<sub>EQ</sub> WHO<sub>i</sub> [CP<sub>U</sub> t<sub>i</sub> [C<sub>U</sub> *did*] [*Mary have tea with t<sub>i</sub>*]]]?

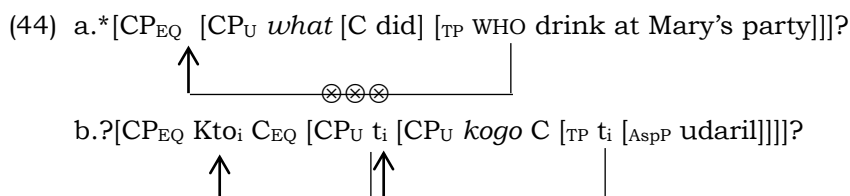
However, as we have seen, complete echo *wh*-extraction out of interrogative contexts is much more restricted and it is subject to parametric variation: such EQs result completely ungrammatical in English (see (37c) and (38c,d)) and Spanish (see (41d)), but they are licit (although slightly deviant) in Russian (see (39d)). This puzzling crosslinguistic variation follows straightforwardly from the current proposal. It is commonly assumed in the literature that, unlike languages of the English

<sup>25</sup> Observe that under this view there is no need to postulate any exceptional nature of such constructions (*contra* Sobin's 2010 pseudo-EQs).



type, Slavic languages, which exhibit obligatory multiple *wh*-fronting in standard, non-echo *wh*-questions, resort to multiple specifiers of CP (see Rudin 1988; Richards 2001; Bošković 2002; among many others).

In EQs based on a previous *wh*-question, the specifier of the CP<sub>U</sub> is occupied by the *wh*-word “inherited” from the utterance, as in (38)-(39), or, in the case of Spanish EQs based on a polar question, (41), this position hosts *si* ‘whether’. However, in Russian EQs there is an additional escape hatch at CP<sub>U</sub>. Thus, it is not surprising that a complete echo *wh*-extraction into CP<sub>EQ</sub>, through the edge of CP<sub>U</sub>, is allowed only in this language. The contrast is schematically represented below, for the English EQ in (38c) and the Russian one in (39d):<sup>26</sup>



In this respect, Spanish is similar to English in that it resorts to single *wh*-movement in true questions with more than one *wh*-word (see below). Hence, this language does not resort to multiple specifiers of CP:

- (45) a. ¿Quién<sub>1</sub> ha visto a quién<sub>2</sub>?  
 who.NOM has seen who.ACC  
 ‘Who has seen whom?’

<sup>26</sup> As for the marginal status of Russian EQs with complete *wh*-movement, (38d), I argue that it can be accounted for in terms of Relativized Minimality (since Rizzi 1990), namely its reformulation in terms of sensitivity to the feature-specification of the involved elements (Starke 2001; see also Rizzi 2013):

(i) In the configuration [...X ...Z ...Y], a local relation cannot hold between X and Y if Z intervenes and Z fully matches the specification of X in the relevant morphosyntactic features [adopted from Rizzi 2013:179].

Roughly, the echo *wh*-word in (43a) can pass over the non-echo *wh*-item because the former is more richly specified (by being [+anaphoric]) than the latter. For a detailed discussion of this issue, see Chernova (2015).

b.\*¿Quién<sub>1</sub> a quién<sub>2</sub> ha visto?<sup>27</sup>  
 who.NOM who.ACC has seen

Thus, the ungrammaticality of Spanish EQs with echo *wh*-movement into the leftmost position, as in (41d), is also expected under the same logic as in (44).

#### 4.2. Partial echo *wh*-movement

Let us consider now partial echo *wh*-movement to some preverbal position, attested in Russian and Spanish EQs (see the examples below):

- (46) Spanish  
 ??¿Que qué QUIÉN<sub>i</sub> ha leído t<sub>i</sub>?  
 that what.ACC who.NOM has read  
 ‘What did WHO read?’
- (47) Russian  
 Kogo KTO<sub>i</sub> udaril t<sub>i</sub>?  
 who.ACC who.NOM hit  
 ‘Whom did WHO hit?’

Recall our discussion that in these languages, in addition to the phase domains projected by C and *v*, there is another intermediate functional projection that can act as a phase and, consequently, can exhibit A-bar properties (and host elements undergoing A-bar movement). In Russian, such phasal properties are assumed by AspP and in Spanish, by TP.

I suggest that in EQs with partial *wh*-extraction, there is an echo Q-particle that is merged *at distance* from the echo *wh*-word and resorts to adjunction. This means that such Q<sub>EQ</sub>

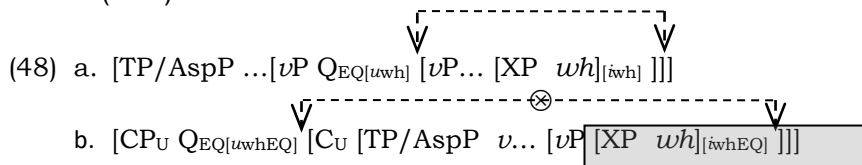
---

<sup>27</sup> It has been argued in the literature that under particular pragmatic contexts Spanish can allow multiple *wh*-fronting (see Etxepare and Uribe-Etxebarria 2005; Uriagereka 2005; Gallego 2017), although, presumably, the lower *wh*-word does not move as high as the first one:

(i) ?No sé quién<sub>1</sub> a quién<sub>2</sub> ha enviado una carta.  
 NEG know who.NOM to who.ACC has sent a letter  
 ‘I don’t know who sent the letter to whom’ (Uriagereka 2005: 2)

does not need to project its  $Q_{EQ}$ , thus it does undergo movement into  $CP_{EQ}$  by itself to check its  $Q_{EQ}$ -feature. Assuming Cable's (2010) Q-based theory, such Q can adjoin low (e.g., within  $\nu P$ ) or high (e.g., at  $CP_U$ ). I assume that such Q bears an unvalued instance of  $[wh_{EQ}]$ ; thus, it has to undergo agreement with the echo *wh*-word, bearer of the valued instance of the matching feature. The latter must be visible to the former to be able to agree.

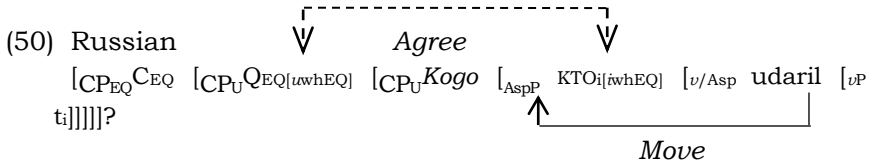
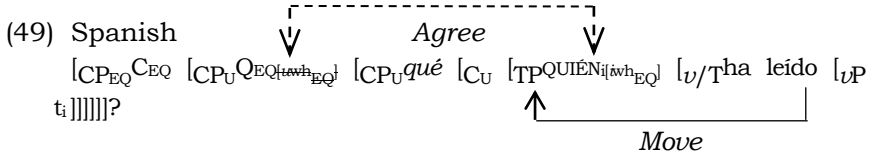
Recall that, in principle, the Q-particle can be merged anywhere in the tree. Suppose that  $Q_{EQ}$  is adjoined low (say, at the edge of  $\nu P$ ), as represented in (48a). From this position, it can agree with the echo *wh*-word, valuing its instance of  $[wh_{EQ}]$ , and then it undergoes successive-cyclic movement into its scope position, the edge of  $CP_{EQ}$ . By being adjoined, such  $Q_{EQ}$  does not pied-pipe the echo *wh*-word, leaving it in-situ. Suppose, however, that the  $Q_{EQ}$  is merged high (say, at  $CP_U$ ), as shown in (48b):



In (48b), the *wh*/Q agreement cannot take place, as the goal (the echo *wh*-word) is within the domain of the extended phase and, hence, it is invisible to the higher probe  $Q_{EQ}$ . Given that the formal imperfection on  $Q_{EQ}$  cannot be deleted, such derivation fails.

The data in (46)-(47) suggest that in Spanish and Russian EQs with partial *wh*-movement the  $Q_{EQ}$  is merged high, as we see that the echo *wh*-word (the goal) raises to a preverbal position. I assume that in these languages, in order to escape the extended phase domain and remain visible to the probe, the echo *wh*-word moves to the edge of TP or AspP, respectively. Notice that the “inherited”, non-echo *wh*-word (at Spec, $CP_U$ ) cannot intervene between the probe and the goal, as it is specified with a different set of features (i.e., it does not bear  $[wh_{EQ}]$ ).

This is illustrated below for the Spanish *wh*-EQ in (46), and the Russian one in (47), respectively:<sup>28</sup>

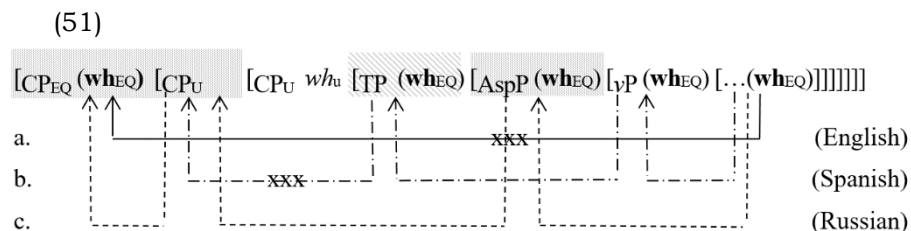


Once agreement takes place and the Q<sub>EQ</sub> deletes its formal imperfection, it undergoes local movement into the edge of CP<sub>EQ</sub>, reaching its scope position.

Evidently, the option of partial *wh*-movement is not available in English, as this language does not resort to the extension of the vP phase. Thus, the echo *wh*-phrase remains low, in-situ.

The successive-cyclic nature of echo *wh*-movement is schematically represented below (the shadowed zones represent the additional host positions and escape hatches for the echo *wh*-word that are available in Russian and Spanish, but absent from English):

<sup>28</sup> A reviewer wonders what happens if the Q-particle is merged at the very end of the derivation, when C<sub>EQ</sub> is in the structure. In fact, under Cable's (2010) Q-theory, the Q-particle can be merged anywhere in the tree from where it c-commands the *wh*-phrase. As we have seen, if the Q is merged vP-internally and projects a QP, it triggers explicit *wh*-movement into the left periphery. However, the Q-particle can also be adjoined directly to C<sub>U</sub> or C<sub>EQ</sub>. In such a case, the Q-particle will not project any QP, binding the *wh*-item at distance; consequently, the *wh*-word will remain in-situ. In principle, I assume that such derivation is possible for EQs with *wh*-in-situ (e.g., *Mary had tea with WHO?*; *Did Mary have tea with WHO?*; *What did WHO drink at Mary's party?*; etc.; for a detailed discussion, see Chernova 2015). In this paper, however, I focus on EQs with explicit movement, especially those contexts in which a QP has to circumvent barriers to reach the CP<sub>EQ</sub>.



To sum up, assuming the particular syntactic structure of *wh*-EQs, the intuition is that echo *wh*-movement, just as standard *wh*-movement, proceeds successive-cyclically through the available escape hatches, and it is subject to certain parametric variation. The  $Q_{EQ}$ -morpheme has to reach its scope position: the edge of the highest phase,  $CP_{EQ}$ . Depending on the merging options of  $Q_{EQ}$ , the echo-introduced *wh*-word can either be pied-piped into the left periphery of the question or undergo partial movement to the edge of lower phases to be visible for the probe  $Q_{EQ}$ .

## 5. Conclusions

In this paper, I addressed the syntax of *wh*-EQs in three typologically different *wh*-fronting languages (English, Spanish and Russian). I argued that echo *wh*-movement is parallel to standard *wh*-movement in true questions and is subject to similar kinds of restrictions. Crucially, echo *wh*-movement also proceeds successive-cyclically, through the available escape hatches on its path. I offered new empirical data showing that in MWF languages (e.g., Russian) the echo *wh*-item can be fronted into the leftmost position of an EQ independently of the clause-type of the echoed utterance. Meanwhile, in languages resorting to single *wh*-fronting in multiple questions (e.g., English and Spanish) the possibility of echo *wh*-movement to the left edge of the question is restricted by the type of the utterance: whether it is declarative or interrogative. In addition, I argued that in Russian and Spanish an echo *wh*-

word may also undergo partial movement to a lower, preverbal, position; this option being also parametrically restricted.

Following Sobin (2010), I assumed that *wh*-EQs have a particular syntactic structure, with two CP levels. As for the nature of the interrogative movement, I assumed the insights of Cable's (2010) Q-based theory and extended it to the derivation of EQs. That is, there is a particular echo Q-morpheme, which is merged in EQs and which regulates the semantics of the echo *wh*-words; during the derivation, the morpheme has to reach its scope position, the specifier of CP<sub>EQ</sub>.

On the one hand, I argued that the Q<sub>EQ</sub>-particle may project a QP<sub>EQ</sub>, which also dominates its sister, the echo *wh*-word, and pied-pipes it into the left periphery of the question. However, to reach the specifier of CP<sub>EQ</sub>, such QP<sub>EQ</sub> has to pass through the lower phase edge, CP<sub>U</sub>. Thus, it is expected that the syntactic character of the echoed utterance (declarative *vs.* interrogative) would restrict the availability of the specifier of CP<sub>U</sub> as an escape hatch. That is, we expect that such movement is allowed in the case of EQs built on a previous declarative utterance, while it is restricted if the echoed utterance is interrogative. In addition, I argued that the typology of *wh*-movement in true multiple questions (i.e., whether a particular language can make use of multiple specifiers of CP) also determines the final derivational outcome, as it may enable additional escape hatches for extraction of the echo QP.

On the other hand, the Q<sub>EQ</sub>-morpheme may also resort to adjunction and be merged at distance from the echo *wh*-word. In such case, the latter cannot be pied-piped into the leftmost position of the clause together with the Q<sub>EQ</sub>; however, the *wh*-item still may undergo raising to an edge of a lower phase to remain visible to its probe. Again, this option is also parametrically restricted. I argued that it relies on the mechanism of extension of the *v*P phase in Spanish and Russian, which has been proposed in the literature on independent grounds.

The two possibilities regarding echo *wh*-movement attested among the languages under consideration naturally follow

from the view offered in this paper. The discussion, hopefully, sheds some more light on the nature of such understudied phenomena as EQs.

## References

- Baker, Carl Lee (1970). "Notes on the description of English questions: the role of an abstract question morpheme". *Foundation of Language* 6: 197-217.
- Bartels, Christine (1997). *Towards a Compositional Interpretation of English Statement and Question Intonation*. PhD dissertation, University of Massachusetts.
- Borik, Olga, Tanya Reinhart (2004). "Telicity and perfectivity: Two independent systems". In: László Hunyadi, Györgi Rákosi, Enikő Tóth (eds.). *Proceedings of LoLa: The Eighth Symposium on Logic and Language*. Debrecen: Hungary, 13-34.
- Borik, Olga (2006). *Aspect and Reference Time*. Oxford: Oxford University Press.
- Bošković, Željko (2002). "On multiple *wh*-fronting". *Linguistic Inquiry* 33: 351-383.
- Bošković, Željko (2014). "Now I'm a phase, now I'm not a phase: on the variability of phases with extraction and ellipsis". *Linguistic Inquiry* 45: 27-89.
- Cable, Seth (2010). *The Grammar of Q: Q-particles, Wh-movement, and Pied-Piping*. Oxford: Oxford University Press.
- Carnie, Andrew (2006). *Syntax: A Generative Introduction*. Oxford: Blackwell.
- Cheng, Lisa, Johan Rooryck (2000). "Licensing *wh*-in-situ". *Syntax* 3: 1-19.
- Cheng, Lisa (1991). *On the Typology of Wh-Questions*. PhD dissertation, MIT.
- Chernova, Ekaterina (2013). "On the syntax of echo *wh*-movement in light of multiple spell out". *Linguistic Analysis* 38: 153-282.
- Chernova, Ekaterina (2015). *The Syntax of Wh-movement in Multiple (True and Echo) Question: A Q-based Approach*. PhD dissertation, Universitat de Girona.

- Chernova, Ekaterina (2017). "On instances of wh-movement in Spanish echo-questions". *Borealis: An International Journal of Hispanic Linguistics* 6: 181-205.
- Chomsky, Noam, Howard Lasnik (1977). "Filters and control". *Linguistic Inquiry* 8: 425-504.
- Chomsky, Noam (1977). "On wh-movement". In: Peter Culicover (ed.). *Formal Syntax*. New York: Academic Press, 71-132.
- Chomsky, Noam (1995). *The Minimalist Program*. Cambridge, MA: MIT Press.
- Chomsky, Noam (2000). "Minimalist inquiries: the framework". In: Roger Martin, David Michaels, Juan Uriagereka, Samuel Keyser (eds.). *Step by Step. Essays on Minimalist Syntax in Honour of Howard Lasnik*. Cambridge, MA: MIT Press, 89-155.
- Chomsky, Noam (2001). "Derivation by phase". In Michael Kenstowicz (ed.). *Ken Hale: A Life in Language*. Cambridge, MA: MIT Press, 1-52.
- Chomsky, Noam (1986). *Barriers*. Cambridge, MA: MIT Press.
- Cooper, Robin (1983). *Quantification and Syntactic Theory*. Dordrecht: Reidel.
- Culicover, Peter (1976). *Syntax*. New York: Academic Press.
- Dalrymple, Mary, Ronald M. Kaplan (2000). "Feature indeterminacy and feature resolution". *Language* 76: 759-798.
- den Dikken, Marcel (2007). "Phase Extension: Contours of a theory of the role of head movement in phrasal extraction". *Theoretical Linguistics* 33: 1-41.
- Dumitrescu, Domnita (1992). "Sintaxis y pragmática de las preguntas cuasi-eco en español". In: Antonio Vilanova (ed.). *Actas del X Congreso de la Asociación Internacional de Hispanistas* (vol. 4). Barcelona: PPU, 1323-1338.
- Dyakonova, Marina (2009). *A Phase-Based Approach to Russian Free Word Order*. Utrecht: LOT Publications.
- Escandell, M. Victoria (1999). "Los enunciados interrogativos: Aspectos semánticos y pragmáticos". In: Ignacio Bosque, Violeta Demonte (eds.). *Gramática Descriptiva de la Lengua Española*. Vol. 3. Madrid: Espasa Calpe, 3929-3991.
- Escandell, M. Victoria (2002). "Echo-syntax and metarepresentations". *Lingua* 112: 871-900.
- Etxepare, Ricardo, Myriam Uribe-Etxebarria (2012). "Las preguntas de qu-in situ en español: Un análisis derivacional". In: José



- M. Brucart, Ángel J. Gallego (eds.). *El movimiento de constituyentes*. Madrid: Visor, 251-272.
- Etxepare, Ricardo, Myriam Uribe-Etxebarria (2015). "Word order in *wh*-questions". Talk given at the III Workshop on Syntax and Semantics (WoSS III). Université de Nantes, 8–9 September.
- Fanselow, Gisbert (2005). "Partial *Wh*-Movement". In: Michael Everaert, Henk van Riemsdijk (eds.). *The Blackwell Companion to Syntax*. Vol. 3. Malden: Blackwell, 436-492.
- Fiengo, Robert (2007). *Asking Questions: Using Meaningful Structures to Imply Ignorance*. Oxford: Oxford University Press.
- Gallego, Ángel (2006). "Phase effects in Iberian Romance". In: Nuria Sagarra, Almeida Jacqueline Toribio (eds.). *Selected Proceedings of the 9th HLS*. Somerville, MA: Cascadilla Proceedings Project, 43-55.
- Gallego, Ángel (2007). *Phase Theory and Parametric Variation*. PhD dissertation, Universitat Autònoma de Barcelona.
- Gallego, Ángel (2010). *Phase Theory*. Amsterdam: John Benjamins.
- Gallego, Ángel (2017). "Multiple *wh*-movement in European Spanish." In Elena Castroviejo, Olga Fernández-Soriano, Isabel Pérez-Jiménez (eds.). *Boundaries, Phases, and Interfaces: Case Studies in Honor of Violeta Demonte*. Amsterdam: John Benjamins, 195-222.
- Ginzburg, Jonathan, Ivan Sag (2000). *Interrogative Investigations*. Stanford, CA: CSLI publications.
- Hagstrom, Paul A. (1998). *Decomposing Questions*. PhD dissertation, MIT.
- Harley, Heidi, Elizabeth Ritter (2002). "Person and number in pronouns: A feature-geometric analysis". *Language* 78: 482-526.
- Hernanz, M. Lluïsa (2012). "Sobre la periferia izquierda y el movimiento: El complementante *si* en español". In: José M. Brucart, Ángel J. Gallego (eds.). *El movimiento de constituyentes*. Madrid: Visor, 151-171.
- Higginbotham, James, Robert May (1981). "Questions, quantifiers and crossing". *Linguistic Review* 1: 41-80.
- Iwata, Seizi (2003). "Echo questions are interrogatives? Another version of a metarepresentational analysis". *Linguistics and Philosophy* 26: 185-254.
- Jiménez, Maria Luisa (1997). *Semantic and Pragmatic Conditions on Word Order in Spanish*. PhD dissertation, Georgetown University.

- Kishimoto, Hideki (2005). "Wh-in-situ and movement in Sinhala questions". *Natural Language and Linguistic Theory* 23: 1-51.
- Kotek, Hadas (2014). "Wh-fronting in a two-probe system". *Natural Language and Linguistic Theory* 32: 1105-1143.
- Krifka, Manfred (2001). "Quantifying into question acts". *Natural Language Semantics* 9: 1-40.
- Noh, Eun-Ju (1998). "Echo questions: Metarepresentations and pragmatic enrichment". *Linguistics and Philosophy* 21: 603-628.
- Parker, Frank, John Pickeral (1985). "Echo questions in English", *American Speech* 60: 337-347.
- Pesetsky, David (1987). "Wh-in-situ: Movement and unselective Binding". In: Eric Reuland, Alice G. B. ter Meulen (eds.). *The Linguistic Representation of (In)definiteness*. Cambridge, MA: MIT Press, 98-129.
- Pesetsky, David (2007). "Property delay (remarks on "phase extension" by Marcel den Dikken)". *Theoretical Linguistics* 33: 105-120.
- Pesetsky, David, Esther Torrego (2001). "T-to-C movement: Causes and consequences". In: Michael Kenstowicz (ed.). *Ken Hale: A Life in Language*. Cambridge, MA: MIT Press, 355-426.
- Pesetsky, David, Esther Torrego (2004). "Tense, case, and the nature of syntactic categories". In: Jacqueline Guéron, Jacqueline Lecarme (eds.). *The Syntax of Time*. Cambridge, MA: MIT Press, 495-537.
- Ramchand, Gillian (2004). "Time and event: The semantics of Russian prefixes". *Nordlyd* 32: 323-261.
- Reglero, Lara (2007). "Wh-interrogatives in Spanish". *Probus* 19: 267-297.
- Richards, Norvin (2001). *Movement in Language: Interactions and Architectures*. Oxford: Oxford University Press.
- Rigau, Gemma (1984). "De com 'si' no és conjunció i d'altres elements interrogatius". *Estudis Gramaticals: Working Papers in Linguistics*. Bellaterra: Universitat Autònoma de Barcelona, 249-278.
- Rizzi, Luigi (1990). *Relativized Minimality*. Cambridge, MA: MIT Press.
- Rizzi, Luigi (2001). "On the position 'interrogative' in the left periphery of the clause". In: Guglielmo Cinque, Giampaolo Salvi (eds.). *Current Studies in Italian Syntax*. Amsterdam: North Holland, 287-296.
- Rizzi, Luigi (2013). "Locality". *Lingua* 130: 169-186.
- Rudin, Catherine (1988). "On multiple questions and multiple wh-fronting". *Natural Language and Linguistic Theory* 6: 445-502.

- Šimík, Radek (2009). "Amazement echo questions as rethoric echo questions". Paper presented in the class of A. Kratzer and L. Selkirk "Meaning and intonation". University of Amherst.
- Sobin, Nicholas (1978). "On echo questions in English". In: Donald Lance, Daniel Gulstad (eds.). *Papers from the 1977 Mid-America Linguistics Conference*. Columbia: University of Missouri, University Extension Division, 247-259.
- Sobin, Nicholas (1990). "On the syntax of English echo questions". *Lingua* 81: 141-167.
- Sobin, Nicholas (2010). "Echo questions in the minimalist program". *Linguistic Inquiry* 41: 131-148.
- Starke, Michal (2001). *Move Dissolves into Merge: A Theory of Locality*. PhD dissertation, University of Geneva.
- Sudo, Yasutada (2007). "A metalinguistic semantics for echo questions". In: Maria Aloni, Paul Dekker, Floris Roelofsen (eds.). *Proceedings of the Sixteenth Amsterdam Colloquium*. Amsterdam: ILLC, University of Amsterdam, 205-211.
- Suñer, Margarita (1991). "Indirect questions and the structure of CP: Some consequences". In: Hector Campos, Fernando Martínez-Gil (eds.). *Current Studies in Spanish Linguistics*. Washington: Georgetown University Press, 283-312.
- Svenonius, Peter (2000). "Quantifier movement". In: Peter Svenonius (ed.). *The Derivation of VO and OV*. Amsterdam: John Benjamins, 255-292.
- Svenonius, Peter (2004a). "Slavic prefixes inside and outside VP". *Nordlyd* 32: 205-251.
- Svenonius, Peter (2004b). "Russian prefixes are phrasal". Ms. University of Tromsø.
- Uriagereka, Juan (1995). "An F position in Western Romance". In: Katalin E. Kiss (ed.). *Discourse Configurational Languages*. New York: Oxford University Press, 153-175.
- Uriagereka, Juan (2005). "A markovian syntax for adjuncts". Talk given at the CLT Seminar, UAB, June 6.
- Uribe-Etxebarria, Myriam (2002). "In situ questions and masked movement". *Linguistic Variation Yearbook* 2: 259-303.
- Vlachos, Christos (2012). *Wh-constructions and the Division of Labour between Syntax and the Interfaces*. PhD dissertation, University of Patras.

- Watanabe, Akira (2001). "Wh-in-situ languages". In: Mark Baltin, Chris Collins (eds.). *The Handbook of Contemporary Syntactic Theory*. Oxford: Blackwell Publishers, 203-225.
- Watanabe, Akira (2002). "Wh and operator constructions in Japanese". *Lingua* 113: 519-558.

### **Acknowledgements**

I thank Ángel Gallego, Avel·lina Suñer, M.<sup>a</sup> Llúisa Hernanz, Josep M.<sup>a</sup> Brucart, Ian Roberts, Theresa Biberauer, Michelle Sheehan, Michael Starke, Ricardo Etxepare, Aritz Irurtzun, Peter Svenonius and Antonio Fábregas for having discussed with me different aspects of this paper. I am also deeply grateful to two anonymous reviewers whose valuable comments and suggestions helped to improve the final version. Finally, I thank all my native informants for their grammatical judgments.

Ekaterina Chernova  
ORCID iD: 0000-0002-5790-6482  
Escola Oficial d'Idiomes de Girona  
Josep Viader i Moliner 16  
17005 Girona  
Spain  
echernov@xtec.cat

## **On complementizers and embedded gapping in English, Spanish and Polish**

JAVIER FERNÁNDEZ-SÁNCHEZ

*Received 15.02.2021,  
received in revised form 03.09.2021,  
accepted 03.09.2021.*

### **Abstract**

This paper examines two sequences which display gapping under two different embedding configurations in English, Spanish and Polish. I claim that the different distribution of the finite complementizer in these configurations and across these three languages provides further evidence for the idea that gapping is not a uniform phenomenon, and that different structures may correlate with different heights at which coordination can take place in gapping.

### **Keywords**

syntax, ellipsis, gapping, complementizers, coordination

### **O spójnikach podrzędnych i podrzędnym *gapping* w języku hiszpańskim, angielskim i polskim**

### **Abstrakt**

W tym artykule przeanalizowano dwie sekwencje, które wykazują *gapping* w dwóch różnych strukturach podrzędnych w języku angielskim, hiszpańskim i polskim. Twierdę, że różny rozkład skończone-

go spójnika w tych konfiguracjach, i w tych trzech językach, dostarcza dalszych dowodów na to, że *gapping* nie jest zjawiskiem jednorodnym i że różne struktury mogą korelować z różnymi wysokościami, na których może mieć miejsce koordynacja w *gapping*.

### Słowa kluczowe

składnia, elipsa, *gapping*, spójniki podrzędne, koordynacja

## 1. Introduction

Gapping is a phenomenon in which the verb in the rightmost conjunct of a sentence coordination structure is elided under identity with the verb in the leftmost conjunct (1a), which I will refer to as the *antecedent clause*. Examples (1b) and (1c) show that ellipsis may target elements other than the main verb, like complements or adjuncts, even if these elements do not appear to conform a constituent (1c):

- (1) a. *Linda studies psychology, and her brother ~~studies~~ biology.*  
 b. *I will travel to Sri Lanka in the summer, and my neighbour ~~will travel to Sri Lanka~~ in autumn.*  
 c. *I will travel to Sri Lanka in the summer, and my neighbour ~~will travel to Israel in the summer~~.*

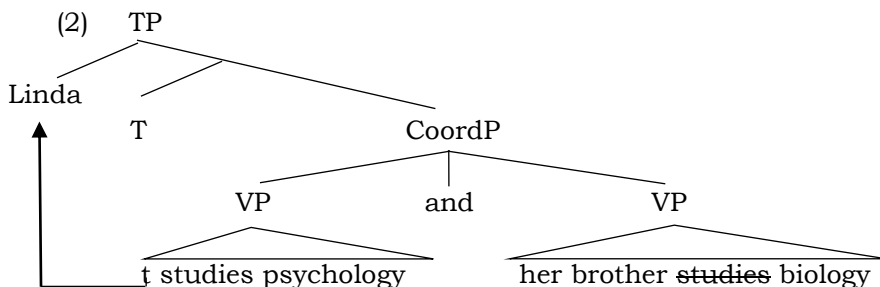
For the purposes of this paper, the examples in (1) will be referred to as canonical gapping, which can be defined as gapping occurring in matrix clauses.<sup>1</sup> Very broadly speaking, the various existing analysis of canonical gapping differ along two main questions: (i) what formal mechanism is responsible for the gap in the second conjunct?; and (ii) at what height does coordination take place in gapping? With respect to the first

---

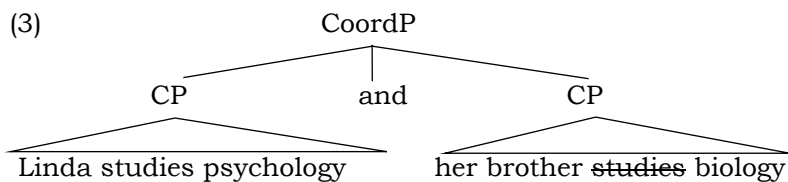
<sup>1</sup> Most of the literature on gapping has indeed focused on canonical gapping. The term is not supposed to have any theoretical relevance, I use it simply to distinguish it from the two gapping structures that I examine in this paper.

question, I will assume that ellipsis involves deletion of syntactic material at PF (i.a. Sag 1976). Following standard practice, I represent elided material in strikethrough text, as illustrated in (1). With respect to the second question, two main analyses have been put forth, which are typically referred to as *low* and *high coordination* accounts.

Low coordination analyses (Coppock 2001, Lin 2002, Johnson 2009, i.a.) posit that coordination in gapping holds at the level of the VP. Under these accounts, the example in (1) would receive the structure in (2). For simplicity reasons, I will represent coordination using non-binary branching, see Zhang (2010) for discussion.<sup>2</sup>

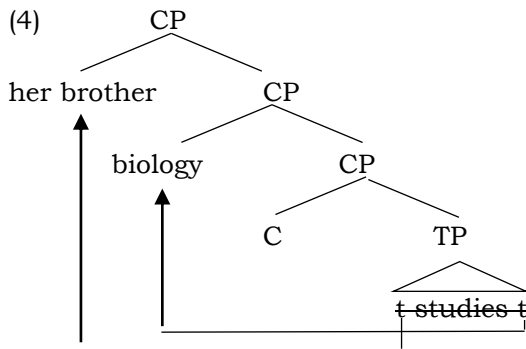


Alternatively, under *high coordination* accounts (Neijt 1979, Hartmann 2000, Reich 2006, *inter alia*), canonical gapping involves coordination of two CPs. Compare (2) to (3):



<sup>2</sup> The representation in (2) is not without its problems. For example, it is unclear why extraction of the preverbal subject from the leftmost VP does not violate the Coordinate Structure Constraint, or how the subject is licensed in the second conjunct; see Johnson (2009) for discussion.

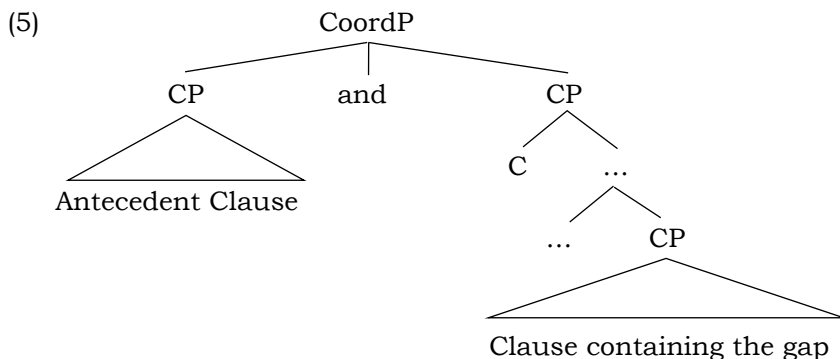
One issue with the representations in (2) and (3) is that the PF deletion operation they display appears to target non-constituents (see the discussion on Fernández-Sánchez 2020: chap.4). To avoid this, it is customary to assume that remnants, i.e. the elements that survive ellipsis – in (1) those would be the DP *her brother* and the NP *biology* – undergo movement to the left edge of the ellipsis domain. Therefore, as an illustration, the rightmost CP in (3) would actually look like (4):<sup>3</sup>



It is interesting to see that canonical gapping is *a priori* compatible with both low and high coordination structures. In this short paper, I focus on two non-canonical gapping configurations in three languages, namely English, Spanish and Polish. These configurations involve gapping in subordination contexts: Non-Canonical Gapping 1 (NCG1) displays an asymmetric coordination structure, where the clause containing the gap is not directly coordinated with its antecedent:

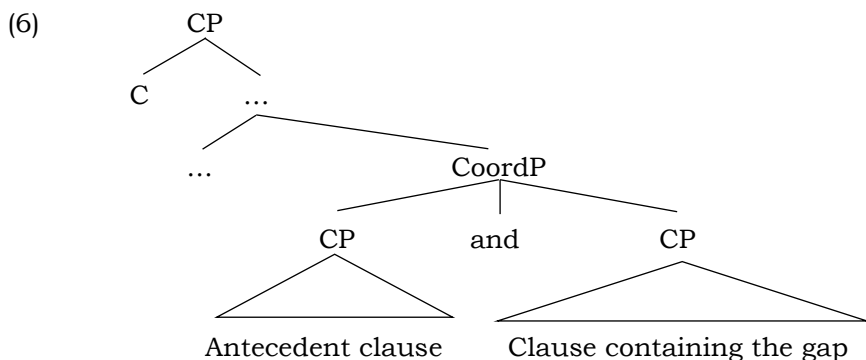
<sup>3</sup> For low coordination accounts, remnants would move to the left edge of the CP. For the purposes of this paper the exact position and motivation for such movements are not relevant. We can assume, following the contrastive nature of the remnants in gapping (Kuno 1976), that they move to the specifier of a Focus Phrase.





Such cases have been argued to be ungrammatical in English (Hankamer 1979), but they have been reported to be fine in Spanish and Polish (Fernández-Sánchez 2016), as well as in English (Wurmbrand 2017) and in other languages like Farsi (Farudi 2013), Georgian and Russian (Erschler 2016). I address NCG1 in section 2. Note that the structure in (5), as opposed to cases of canonical gapping, is absolutely incompatible with a low coordination structure, and must be given a high/clausal coordination analysis.

In turn, Non-Canonical Gapping 2 (NCG2) involves cases where the clause containing the gap is directly coordinated with its antecedent clause, just like in canonical gapping (cf. 1); however, in this case, the entire coordination is embedded under one main verb. NCG2 is illustrated in (6):



While (6) is indeed compatible theoretically with both a high and a low coordination account, we will see in section 3 that there are reasons to believe that NCG2 involves a low coordination structure, which means that the representation in (6) will not be entirely accurate. The claims in this paper suggest that gapping is therefore not a unified phenomenon, a conclusion which goes in line with previous research (Repp 2009, Centeno 2011, Jung 2016...).

Before concluding the paper, in section 4 I will tentatively address the syntax of an understudied gapping string which I will take to be a run-of-the-mill case of NCG1 where the embedding predicate is in turn gapped.

## 2. Non-Canonical Gapping 1

### 2.1. The No Embedded Constraint

Hankamer (1979) proposed that gapping was subject to the *No Embedded Constraint* (NEC hereafter), which essentially states that neither gaps (7a) nor their antecedents (7b) can be embedded (examples from Hankamer):

- (7) a. \* [*Alfonse stole the emeralds*] and [*I think [that Mugsy ~~stole~~ the pearls]*].  
 b. \* [*I think [that Alfonse stole the emeralds] and [Mugsy ~~stole~~ the pearls]*].<sup>4</sup>

In this paper I have nothing to say about (7b), see Toosarvandani (2016). With respect to (7a), low coordination accounts to gapping, *cf.* (2), can easily explain this restriction: one single T head cannot be shared by two VPs if one of them is embedded in another T head. However, equivalent sentences

---

<sup>4</sup> It is important to note that (7b) is grammatical under the reading where coordination holds at the level of the embedded clause. This configuration, illustrated in (i), corresponds to what I call NCG2 in this paper; see section 3 for details:

(i) I think that [*Alfonse stole the emeralds*] and [*Mugsy ~~stole~~ the pearls*].

to (7a) have been argued to exist in other languages like Polish (8) or Spanish (9):

- (8) *Lukasz pojechał do Tajlandii, a zgaduję, że jego brat*  
 Łukasz travelled to Thailand and guess that his brother  
*do Berlina.*  
 to Berlin  
 'Łukasz travelled to Thailand and I guess that his brother  
 travelled to Berlin.'
- (9) *Susana compró una casa en el centro de Madrid y*  
 Susana bought a house in the centre of Madrid and  
*diría que Martina un apartamento en la playa.*  
 would say that Martina an apartment in the beach  
 'Susana bought a house in the centre of Madrid, and I'd say  
 that Martina bought an apartment by the beach.'

Note that data like (8) or (9) can only be accounted for under a high coordination analysis with clausal ellipsis applying in the embedded clause.<sup>5</sup> The question is: why would English be different from these languages? Is this a typological split? It is important to mention, however, that English is not that different from Spanish or Polish, despite Hankamer's initial observation: structures like (7a) are possible provided that, as observed by Wurmbrand (2017), no complementizer precedes the remnants:

- (10) Alfonso stole the emeralds and I think Mugsy the pearls.<sup>6</sup>

---

<sup>5</sup> Of course this does not mean that gapping in these languages must *always* involve high coordination structures. As an anonymous reviewer mentioned, various authors have developed eclectic accounts of gapping where both high and low coordinations are involved in different gapping strings within the same language; see Repp (2009), Centeno (2011) or Wong (2016); The main claim in this paper is, precisely, that the two configurations under scrutiny here must involve different coordination heights.

<sup>6</sup> An anonymous reviewer wonders whether this is truly a case of embedding, or whether (10) involves a run-of-the-mill gapping structure where the antecedent clause and the clause containing the gap are directly coordinated and the sequence "I think" is a parenthetical comment clause (Schneider 2007, Griffiths 2013) which provides an epistemic/evidential qualification over a proposition. First, the equivalent structures in Polish and Spanish

In order to capture the data, Wurmbrand proposes the following condition:

(11) *The Embedded Gapping Constraint*

Gapping of embedded clauses is only possible when the embedded clause lacks a CP.

To explain the ungrammaticality of (7a) and the grammaticality of (10), she makes the following assumptions: first, she argues – in line with others (Gallego 2009, Bošković 2014, Aelbrecht 2016) that ellipsis is licensed by phasal heads. Second, she contends that, while there are two phasal domains – thematic and propositional, which roughly correspond to  $vP/VP$  and CP respectively – phases should be defined contextually or configurationally. In particular, she defends that phases are the highest head in a phasal domain. Third, she assumes that remnants move to a functional projection (FP) above TP prior to clausal ellipsis, along the lines of (4). Finally, and crucially, she follows Bošković (1997) in claiming that *that*-less embedded clauses are TPs.

After having established the main features of Wurmbrand's analysis, let us see how she derives the facts. Take the example in (10): the verb *think* selects for a TP (following her last premise), as illustrated in (12a). In order for clausal ellipsis to apply, remnants move to a FP above the TP to escape the domain of ellipsis. Ellipsis is then licensed by the highest head in the embedded, propositional phase, which in this case is the head of FP, which triggers ellipsis of its complement, i.e. the TP:

(12) a. *Alfonse stole the emeralds and I think* [<sub>TP</sub> *Mugsy stole the pearls*].

b. ... *and I think* [<sub>FP</sub> *Mugsy*<sub>i</sub> [<sub>FP</sub> *the pearls*<sub>S<sub>k</sub></sub> [<sub>TP</sub> ~~*t<sub>i</sub> stole t<sub>k</sub>*~~ ]]]].

---

are *bona fide* cases of embedding, as evidenced by the overt complementizer, so one would expect embedding to be possible in English as well. Second, regular fragment answers, which display a very similar syntax to gapping (Reich 2006), can be truly embedded (see Weir 2014), see section 2.

If the complementizer is present, as in (7a), then the verb *think* selects for a CP complement. In this scenario, it is C and not F that is the highest phase in the propositional domain. Consequently, C ought to trigger ellipsis of its complement, which encompasses FP. Under this configuration, remnants would stay trapped within the ellipsis spell-out domain.

Although it is an interesting proposal, Wurmbrand's analysis falls short of empirical coverage as it cannot explain why in languages like Spanish or Polish, the complementizer *must* be present; compare (13) to (8) and (9):

- (13) a. \* *Lukasz pojechał do Tajlandii, a zgaduje, jego brat do Berlina.*  
 b. \* *Susana compró una casa en el centro de Madrid, y diría que Martina, un piso en la playa.*

In what follows I claim that NCG1 should be viewed as cases of (embedded) fragment answers, in the sense of Merchant (2004).

## 2.2. Embedded fragments

A question like (14) can be answered, at least, in two ways: one involves repetition of the presupposed content (14a), and the other one involves pronouncing only the focus of the sentence (14b). The latter is what is commonly referred to as a fragment answer:

- (14) Who did you see yesterday?  
 a. Yesterday I saw Mary.  
 b. Mary.

We follow Merchant (2004)'s standard analysis that (14b) is derived from (14a) via clausal ellipsis.<sup>7</sup> In particular, this au-

---

<sup>7</sup> That fragments have an underlying clausal structure can be easily shown in languages with case marking on nominal categories. The equivalent example to (14b) in Polish would be *Marię* (Mary.ACC). The case marking

thor claims that the fragment undergoes movement to a functional projection above the TP prior to ellipsis:

(15) [FP [NP Mary]<sub>i</sub> F [TP ~~yesterday I saw t<sub>i</sub>~~ ] ].

Importantly for the purposes of this paper, fragments can be embedded, as in (16), from Weir (2014: 221); see fn.6:

(16) A: Who is responsible for the 9/11 attacks?  
B: Well, Michael Moore believes Bush.

What I defend here is that NCG1 can be derived by means of the same mechanism that derives (embedded) fragment answers (16). The difference would be that in NCG1 two remnants undergo movement to FP. This analysis is defended on the basis of two parallelisms between embedded fragments and NCG1: (i) the types of predicates under which the remnants can be embedded, section 2.2.1; and (ii) the presence/absence of the complementizer in various languages, section 2.2.2.

### 2.2.1. Embedding predicates

While fragment answers can be embedded, it has noted that not all predicates can embed them (de Cuba and MacDonald 2013, Weir 2014). This is illustrated in these minimal pairs:

(17) A: Who stole the jewels?  
B: I {guess/suppose/think} your son.  
(18) A: Who stole the jewels?  
B: \* I {know/regret/hate} my son.

Let us assume that the key component here is factivity:<sup>8</sup> factive predicates disallow embedded fragments. One possible

---

follows naturally from the fact that (14b) contains an elided verb that assigns accusative to the object.

<sup>8</sup> De Cuba and MacDonald (2013) actually claim that it is not factivity that is at stake, but rather the related – yet independently motivated – notion

explanation is that this is due to the fact that these predicates select for a truncated clausal structure (Vikner 1995, Haegeman 2006) which crucially lacks structural space for remnants to move to prior to ellipsis at PF. The explanation is indeed reminiscent, and correlates nicely, with the classic findings in Hooper and Thompson (1973), who noted that certain syntactic operations like topicalization cannot target the left periphery of clausal complements to factive predicates:

- (19) a. *The inspector explained that each part he had examined carefully.* (Hooper and Thompson 1973: 474, their (50))  
 b. \* *I resent the fact that each part he had to examine carefully.* (ibid.: 479, their (109)).

If NCG1 involves the same structure as embedded fragment answers, we should expect the same restrictions observed in (17) and (18). The following examples show that this prediction is borne out: (20) illustrates that non-factives (a) are compatible with NCG1 in Spanish, and factives (b) are ungrammatical. (21) showcases the same contrast in Polish:

- (20) a. *Alfonso robó las esmeraldas* { *creo/ imagino/...* }  
 Alfonso stole the emeralds and think imagine  
*que Mugsy las perlas.*  
 that Mugsy the pearls  
 b. \* *Alfonso robó las esmeraldas* { *lamento/ odio/...* }  
 Alfonso stole the emeralds and regret hate  
*que Mugsy las perlas*  
 that Mugsy the pearls  
 (21) a. *Alfons ukradł szmaragdów a* { *myślę/ zgduję/...* }, *że*  
 Alfons stole emeralds and think suppose that  
*Mugsy perły.*  
 Mugsy pearls

---

of *referentiality*. For the purposes of this paper factivity is enough, as we are interested simply in the descriptive parallelisms between embedded fragments and NCG1; but see de Cuba and Macdonald (2013) for discussion and references.

- b. \* *Alfons ukradł szmaragdów a {zaluje/wiem/...}, że*  
 Alfons stole emeralds and regret know that  
*Mugsy perły.*  
 Mugsy pearls

### 2.2.2. Presence/absence of the complementizer

Languages differ with respect to whether embedded fragments are preceded by an overt complementizer. English fragment answers cannot be headed by a complementizer (see Fernández-Sánchez and Llinàs-Grau 2017 for discussion), but in Spanish or Polish the complementizer is compulsory:<sup>9</sup>

- (22) A: *What exotic fruit did John buy?*  
 B: *I {guess/think/suppose...} (\*that) a kiwano.*
- (23) A: *¿Qué fruta exótica compró Juan?*  
 B: *{Creo/pienso/supongo...} \*(que) un kiwano.*
- (24) A: *Które owoce egzotyczne kupił Janek?*  
 B: *{Myślę/przypuszczam/zgaduję...}, \*(że) kiwano.*

The distribution of the complementizer in embedded fragments corresponds crosslinguistically with the distribution of the finite complementizer in NCG1 which, taken along with the facts about embedding, strongly suggest that we are indeed dealing with the same phenomenon.

---

<sup>9</sup> In fact, crosslinguistically speaking, languages appear to choose one or the other option, i.e. either obligatory presence of C (Spanish, Catalan, Polish, Czech...) or obligatory absence of C (English, Greek, Dutch...). Trying to relate the obligatory absence of C in English to a *that*-trace effect – which is an environment in which English forces an empty complementizer – is not a fruitful line of research (see Weir 2014: 221-233). Furthermore, Greek behaves like English in forcing an empty complementizer but there is no *that*-trace effect in this language (I am indebted to Anna Roussou for discussion on the Greek data).



### 3. Non-Canonical Gapping 2

The second embedded gapping string I would like to examine involves cases where both the antecedent and clause containing the gap are coordinated at the same level, and coordination appears embedded under a matrix verb. One example is provided in (25):

- (25) I think [Alfonse stole the emeralds] and [Mugsy the pearls].

Contrary to what happens in NCG1, where only a high coordination account is able to explain the data, NCG2 is in principle compatible with both a high and a low coordination analysis (just like any other case of canonical gapping). However, closer scrutiny reveals that a low coordination account fares better with the data.

#### 3.1. Embedding predicates

Suppose that the predicate under which coordination is embedded is a factive one. If NCG2 involved clausal ellipsis like NCG1, then we would expect gapping to be unavailable, given that the coordinated clausal complement would lack the relevant projections for remnants to move to. However, gapping in such cases is possible even with factive predicates, as shown in (26) through (28) for English, Spanish and Polish:

- (26) *I {dislike/regret...} that John goes out with Sonja and Jason  
~~goes out~~ with Lilly.*
- (27) *Me desagrada que Pedro me haya servido la sopa  
to displeases that Pedro to me has served the soup  
fría y su mujer ~~me haya servido~~ el helado derretido.  
cold and his wife the ice-cream melted  
'It displeases me that Pedro has served me the soup cold  
and his wife the ice-cream melted.'*
- (28) *Co za nudne lato tu w Warszawie bez moich  
what for boring summer here in Warsaw without my*

*przyjaciół. Nienawidzę, że Łukasz pojechał do Tajlandii*  
 friends hate that Łukasz travelled to Thailand  
*i Krzys̄ pojechał do Berlina.*  
 and Krzys̄ to Berlin

'What a boring summer here in Warsaw without my friends. I hate it that Łukasz has gone to Thailand and Krzys̄ to Berlin.'

### 3.2. Absence/presence of the complementizer

Hartmann (2001:157) pointed out that in sequences like the one we are dealing with, i.e. NCG2, *that* must be absent in English, an observation she attributes to Fiengo (1974):

- (29) Jim said that Alan went to the ballgame and (\*that) Betsy ~~went~~ to the movies.

In NCG1, the lack of an overt complementizer in English was associated with whatever mechanism disallowed complementizers in embedded fragment answers. The lack of the complementizer in sequences like (29), however, cannot be attributed to that same mechanism, for the simple reason that if a unified account was to be pursued, we would expect the complementizer in Spanish and Polish to be mandatorily overt. This prediction, however, is not borne out: NCG2 must involve a null complementizer in these languages as well:

- (30) a. *Przypuszczam, że Łukasz kupił stary samochód, a*  
 suppose that Łukasz boughtold car and  
 (\* że) *Maciek rower.*  
 thatMaciek bicycle  
 'I guess that Łukasz bought an old car and (\*that) Maciek a bike.'
- b. *Co za nudne lato tu w Warszawie bez moich przyjaciół!*  
*Nienawidzę, że Łukasz pojechał do Tajlandii, a (\*że)*  
*Krzys̄ pojechał do Berlina. (cf.28)*
- (31) a. *Supongo que María traerá las bebidas y (\* que)*  
 suppose thatMaría will bring the drinks and that

*Pedro el postre.*

Pedro the dessert

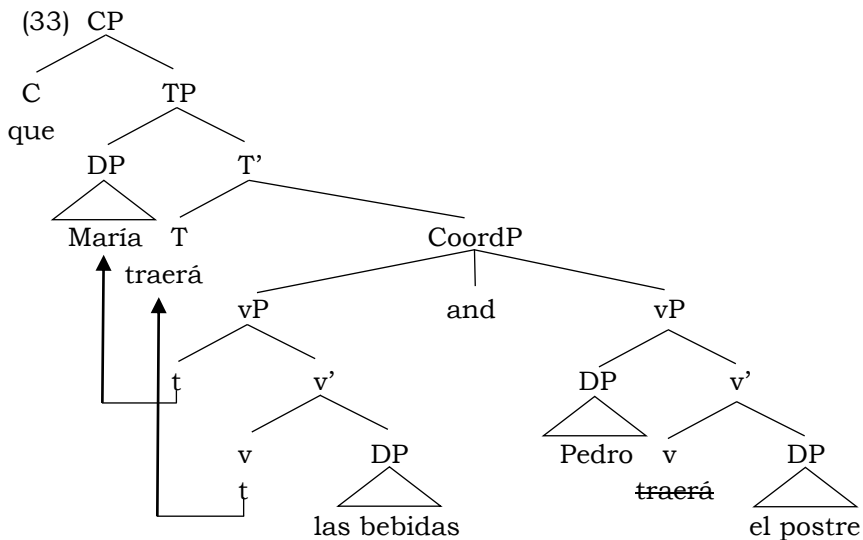
‘I suppose that Mary will bring the drinks and (\* that) Pedro the dessert.’

- b. *Me desagrada que Pedro me haya servido la sopa fría y (\*que) su mujer ~~me haya servido~~ el helado derretido.* (cf.27)

The fact that in NCG2 is incompatible with the complementizer appears to hold for many languages. Hartmann (2001: 158) observes that the same is true in German:

- (32) *Ich glaube, dass Peter mit seiner Frau nach Indien reist und (\*dass) Martin mit seinen Kollegen in die Schweiz.*  
 I think that Peter withhis wife to India travels  
 and that Martin withhis colleagues in the Switzerland  
 ‘I think that Peter travels to India with his wife and Martin ~~travels~~ with his colleagues to Switzerland.’

Taken together, the facts presented in sections 3.1 and 3.2 naturally follow if we assume a low coordination to gapping: take (31a) as an example. According to my proposal, it would involve a structure along the following lines (I only represent the embedded sentence for the sake of simplicity):



In essence, the lack of a complementizer follows obviously from the fact that the second conjunct is not clausal, but rather a vP (but see below). The insensitivity to the factivity of the embedding predicate is expected: under a low coordination account, it is irrelevant whether the left periphery of the embedded predicate is truncated or not. This is so because, again, coordination takes place at a lower level, so no C-domain is involved.

As we saw before, low coordination accounts of gapping assume that coordination holds at the level of the vP. However, as correctly pointed out by a reviewer, the facts presented in this section could still follow from IP-coordination, a solution indeed entertained, but ultimately rejected, by Hartmann (2001) for German. Determining the actual syntactic node at which coordination takes place in NCG2 deserves a more careful examination of the data, a task I leave for further research.

The question that remains is, of course, what is it that bans coordination of two CPs in NCG2. The same reviewer argues that coordination of CPs must be allowed in NCG2 in languages like Spanish at least, because these strings are compatible with gapping involving left dislocated remnants (underlined for expository purposes):

- (34) *Juan aseguró que el dinero lo había guardado en el  
 Juan claimed that the money it had saved in the  
 banco y las joyas en la caja fuerte.  
 bank and the jewels in the strongbox  
 ‘Juan claimed that the money, he had saved it in the bank,  
 and the jewels in the strongbox.’*

Note that under the assumption that left dislocated phrases are in the left periphery of the clause, the DP *las joyas* (‘the jewels’) must be in a CP-position. Data like (34), however, should be handled with care. To start with, note that the second conjunct is not – and in fact it cannot – be headed by

a complementizer, contrary to what would happen if ellipsis had not applied:

- (35) a. *Juan aseguró que el dinero lo había guardado en el banco y (\*que) las joyas en la caja fuerte.*  
 b. *Juan aseguró que el dinero lo había guardado en el banco y \*(que) las joyas las había guardado en la caja fuerte.*

Testing structure with clitic left dislocation is complicated by the fact that, as shown in Fernández-Sánchez (2017), clitic left dislocated phrases often appear in syntactic contexts where it can be shown independently that there is no structural space, a fact that some authors have taken to mean that dislocated phrases should be viewed as parenthetical elements (Fernández-Sánchez 2017, 2020, Fernández-Sánchez and Ott 2020).<sup>10</sup>

But leaving these issues aside, note that there is an important asymmetry between NCG1 and NCG2, in that while it is true that the latter may in theory be compatible with two different structures (whatever they are exactly), the former is not: such cases must necessarily involve a clausal coordination. Given this, we could hypothesize that in cases where two potential derivations would yield the same output, the simplest/most economic one is preferred. Such an economy constraint would be similar to Bošković (1997)'s Minimal Structure Principle:<sup>11</sup>

(36) *The Minimal Structure Principle*

Provided that lexical requirements of relevant elements are satisfied, if two representations have the same lexical structure and serve the same function, then the representation

---

<sup>10</sup> One could still adopt a less radical view and claim, along the lines of Jiménez-Fernández and Miyagawa (2014), that clitic left dislocation involves IP-adjunction. This way, (34) would still be compatible with a lower-than-C coordination. Again, I leave this for further research. What is important is to stress that NCG2 cannot involve CP-coordination.

<sup>11</sup> A very similar conclusion was reached in Fernández-Sánchez (2020), where I looked at the interplay between coordination and ellipsis in dislocation structures.

that has fewer projections is to be chosen as the syntactic representation serving that function.

Similar claims have been made in the generative literature (see Collins (2001) and Dalrymple *et al.* (2015) for discussion. Unfortunately, exploring this falls outside the scope and goals of this paper, so I leave this issue for further research.

#### 4. Double gaps

##### 4.1. Canonical gapping + NCG1

Before concluding the paper, I would like to bring to the fore a construction which, to my knowledge, was firstly noted in Brucart (1987)'s seminal work on ellipsis in Spanish and which involves two verbal gaps being separated by the finite complementizer *que* ('that'):

- (37) *Pedro aseguró que nevaría en los Alpes, y Juan*  
 Pedro claimed that would snow in the Alps and Juan  
 \_\_\_ *que* \_\_\_ *en los Pirineos.*  
 that in the Pyrenees  
 'Pedro claimed that it would snow in the Alps, and Juan  
 claimed that it would snow in the Pyrenees.'
- (38) *Juan confirmó que Susana llegará en avión y Pedro*  
 Juan confirmed that Susana will arrive in plane and Pedro  
 \_\_\_ *que* \_\_\_ *en coche.*  
 that in car  
 'Juan confirmed that Susana will arrive by plane and Pedro  
 confirmed that she would arrive by car.'

Polish allows this construction as well, but English does not:

- (39) \* *John claimed that Susan would travel by plane and Peter*  
 \_\_\_ *that* \_\_\_ *by car.*
- (40) *Janek powiedział, że Andrzej studiował matematykę a*  
 Janek said that Andrzej studied maths and

*Wojciech* \_\_\_ *że* \_\_\_ *inżynierię*.  
*Wojciech*     *that*    *engineering*  
 ‘Janek said that Andrzej studied maths, and Wojciech said  
 that he studied engineering.’

Brucart (1987) contends that the two gaps are the result of the same operation, i.e. gapping. He attributes the unavailability of this construction in English to the fact that the rightmost gap is actually a complex object formed by the unpronounced verb preceded by a null *pro*. Given that English lacks *pro*, the ungrammaticality of (39) follows. Brucart’s explanation would also account for the grammaticality of (40), given that *pro* is available in the grammar of Polish.

The reason to postulate the existence of *pro* comes from Jackendoff (1971)’s suggestion that gaps must contain remnant material at their left and right edges. However, it is well known that remnants of gapping must be focused constituents (Kuno 1976, i.a.) and it is unclear how *pro* can be a focused element. Further, note that under the assumption that the structure under scrutiny is unavailable in English because of the lack of *pro* in this language, we expect this construction to be possible if an overt subject is placed. The prediction, however, is not borne out:

(41) \* *John claimed that Susan would arrive by plane and Pedro*  
       \_\_\_ *that Laura* \_\_\_ *by car*.

I would like to suggest an alternative account of these facts. Descriptively, these examples featuring a double gap can be explained in the following way: the leftmost gap is an instance of canonical gapping – the matrix verb is deleted under identity with the matrix verb in the antecedent clause. The rightmost gap is embedded under the gapped main verb so, in other words, the rightmost gap is an instance of NCG1. There are reasons to believe this. For example, if we try to use a factive

verb as an embedding predicate, the sentence becomes ungrammatical:

- (42) \**Juan lamenta que el gobierno haya subido el IVA y Pedro lamenta que el gobierno haya subido el impuesto de sucesiones.*  
 Juan regrets that the government has raised the VAT and Pedro that the tax of succession  
 ‘Juan regrets that the government has raised VAT and Pedro (regrets) that (the government has raised) the estate tax.’

The ungrammaticality of (42) must be attributed to the right-most gap. We can see this because the two gaps are independent of each other. (43b) is thus ungrammatical for the same reason that (20b) is:

- (43) a. *Juan lamenta que el gobierno haya subido el IVA y Pedro lamenta que el gobierno haya subido el impuesto de sucesiones.*  
 b. \**Juan lamenta que el gobierno haya subido el IVA y Pedro lamenta que el gobierno haya subido el impuesto de sucesiones.*

The question that remains to be addressed is how is it that English disallows this double gap construction. I discuss this in the next section, where I argue that it is the lack of an overt complementizer heading NCG1 in this language that explains the unavailability of double gaps.

#### **4.2. The clause-mate condition on gapping**

To fully understand why English does not allow this construction, it is important that we introduce one locality condition to which gapping is subject: the clause-mate condition on remnants. Empirically, this condition captures the fact that the



gap in (44), here indicated with *e*, can only be interpreted as (45a) and not as (45b):<sup>12</sup>

- (44) *Julia said that Rose speaks Russian and Matthew [e] Polish.*  
 (45) a. [*e*] = speaks *Embedded reading*  
 b. \* [*e*] = said that Rose speaks *Matrix reading*

The explanation for the clause-mate condition cannot simply be to assume that the gap is restricted to only one instance of lexical verb. Ross (1970) already noted that the gap can contain more than one verb (46). In light of data like this one, the relevant generalization is that the gap cannot contain a finite clause boundary:

- (46) a. *I want to try to begin to write a novel, and you a play.*  
 b. *...and you want to try to begin to write a play.*

The clause-mate condition appears to hold crosslinguistically. (47) shows that gapping in Spanish cannot contain a finite clause boundary, whereas (48) illustrates that it may contain a non-finite clausal node. Examples (49) and (50) illustrate the same point with Polish data:

- (47) *Juan aseguró que Susana llegaría en avión y Pedro*  
 Juan claimed that Susana would arrive in plane and Pedro  
 [*e*] *en coche*  
 in car  
 ‘Juan claimed that Susan would arrive by plane and Pedro by car.’  
 a. [*e*] = ... y Pedro ~~llegaría~~ en coche.  
 b. \* [*e*] = ... y Pedro ~~aseguró que Susana llegaría~~ en coche.  
 (48) a. *Luis prometió casarse en Barcelona y Ana [e] en*  
 Luis promised to marry in Barcelona and Ana in in

<sup>12</sup> The labels *matrix* and *embedded* reading capture the height at which coordination must take place in order to derive the corresponding meanings. Therefore, the embedded reading is obtained by coordination at the level of the embedded clause, and the matrix reading by coordination at the root.

*Bilbao.*

Bilbao

‘Luis promised to get married in Barcelona and Ana in Bilbao.’

b. [e] = (prometió) casarse.

(49) *Janek powiedział, że Andrzej studiował matematykę a*

Janek said that Andrzej studied maths and  
*Wojciech [e] inżynierię.*

Wojciech engineering

‘Janek said that Andrzej studied maths and Wojciech engineering.’

a. [e] = ... a Wojciech studiował inżynierię.

b. \*[e] = ... a Wojciech ~~powiedział, że Andrzej studiował~~  
inżynierię.

(50) a. *Janek chce studiować matematykę, a Andrzej [e]*

Janek wants to study maths and Andrzej  
*inżynierię.*

engineering

b. [e] = (chce) studiować.

The clause-mate condition poses a challenging theoretical question, given that aside from gapping, it has been argued to hold in many phenomena which involve ellipsis to the exception of more than one remnant like pseudogapping (Jayaseelan 1990), multiple sluicing (Lasnik 2014) or wh-stripping (Ortega-Santos, Yoshida and Nakao 2014), which strongly suggests that there must be a general, across-construction explanation.

Suppose now that we want to derive a double gap structure in English (51). The matrix verb can undergo ellipsis via canonical gapping (51a). This operation leaves the subject DP remnant and the clausal remnant. Now to derive the embedded gap (NCG1), the remnants-to-be need to undergo movement to the left edge of that embedded clause (Merchant 2014), as shown in (51b). Crucially, as we have seen before, embedded fragments are never preceded by the complementizer in English. In the absence of a complementizer (51c), the two remnants must be interpreted as clause mates, *as per* the clause-mate condition on remnants. In other words: the se-

quence *and Rose [e] Hebrew*, can only be interpreted as *and Rose speaks Hebrew*, and not as *and Rose claimed that Susan speaks Hebrew*.

(51) *John claimed that Susan speaks Arabic and Rose claimed that Susan speaks Hebrew.*

a. Matrix coordination, canonical gapping:

*John claimed that Susan speaks Arabic, and Rose ~~claimed~~ that Susan speaks Hebrew.*

b. TP ellipsis in the clausal remnant in (51b):

*[Hebrew]<sub>i</sub>; that Susan speaks ~~t<sub>i</sub>~~.*

c. Resulting string

*John claimed that Susan speaks Arabic and Rose Hebrew.*

I would like to suggest, thus, that the availability of the double gap construction depends on whether in a particular language embedded fragments (and by extension NCG1) are preceded by an overt complementizer. If they are not, the clause-mate condition on remnants will disallow the intended meaning.

## 5. Conclusions

In this paper I have looked at two structures which involve non-canonical gapping. NCG1, once (wrongly) thought to be ungrammatical (Hankamer 1979) at least in English, must involve clausal coordination, so it is incompatible with low coordination accounts to gapping. In this configuration, the remnants must be headed by an overt complementizer in Spanish and Polish, but in English this complementizer must be empty. Focusing on the English data, Wurmbrand (2017) proposes an account based on a flexible theory of phases, but her analysis is incompatible with the Spanish and Polish data. I have argued instead that the distribution of the finite complementizer in these three languages can be explained if we posit that the mechanism deriving NCG1 is the same one that yields embedded fragment answers. This allows, in turn, to explain why NCG1 is sensitive to the type of embedding predicate.

With respect to NCG2, I have suggested that coordination must be lower than in NCG1. This conclusion is based on the fact that remnants in this configuration are never introduced by a complementizer, even in languages where the complementizer obligatorily heads remnants, as well as by the insensitivity of NCG2 to the type of embedding predicate.

Consequently, this paper shows, in line with others (Repp 2009, Centeno 2011, Jung 2016...) that gapping is not a unified phenomenon, and that this phenomenon can result from the interplay between ellipsis and coordination at different points in the structure.

Finally, I have briefly addressed the syntax of a construction which features two gaps, which are separated by the finite complementizer, in Spanish and in Polish. I have defended that while the rightmost gap is the result of canonical gapping, the embedded gap is an instance of NCG1. This construction does not exist in English for the simple reason that in this language NCG1 cannot be headed by a complementizer, and therefore the two gaps will end up creating a complex string that is not possible to interpret.

## References

- Aelbrecht, Lobke (2016). "What ellipsis can do for phases and what it can't, but not how". *The Linguistic Review* 33: 453-482.
- Bošković, Željko (1997). *The Syntax of Nonfinite Complementation: An Economy Approach*. Cambridge, MA: MIT Press.
- Bošković, Željko (2004). "Now I'm a phase, now I'm not a phase: On the variability of phases with extraction and ellipsis". *Linguistic Inquiry* 45: 27-89.
- Brucart, Josep M. (1987). *La Elisión Sintáctica en Español*. Bellaterra: Publicacions de la UAB.
- Centeno, Naiara (2012). *Gapping and Determiner Sharing in Spanish*. PhD dissertation, UPV/EHU.

- Collins, Chris (2001). "Economy conditions in syntax". In: Mark Baltin, Chris Collins (eds.). *The Handbook of Contemporary Syntactic Theory*. London: Blackwell, 45-61.
- Coppock, Elizabeth (2001). "Gapping: in defense of deletion". In: Mary Andronis, Christopher Ball, Heidi Elston, Sylvain Neuvel (eds.). *Proceedings of Chicago Linguistics Society CLS 37*. Chicago: University of Chicago, CLS, pp.133-148.
- de Cuba, Carlos, Jonathan E. MacDonald (2013). "Referentiality in Spanish CPs". In: Victoria Camacho-Taboada, Ángel J. Jiménez-Fernández, Javier Martín-González, Mariano Reyes-Tejedor (eds.). *Information Structure and Agreement*. Amsterdam: John Benjamins. 117-140.
- Dalrymple, Mary, Ronald M. Kaplan, Tracy Holloway King (2015). "Economy of expression as a principle of syntax". *Journal of Language Modelling* 3: 377-412.
- Erschler, David (2016). "On high and low licensing of gapping". Paper presented at the *Ellipsis Across Borders* conference. University of Sarajevo, Bosnia and Herzegovina.
- Fernández-Sánchez, Javier (2016). "Embedded gaps as embedded fragment answers". Paper presented at the *Ellipsis Across Borders* conference. University of Sarajevo, Bosnia and Herzegovina.
- Fernández-Sánchez, Javier (2017). "Topics at the left edge of infinitive clauses in Spanish and Catalan". *Borealis: an International Journal of Hispanic Linguistics* 5: 111-134.
- Fernández-Sánchez, Javier (2020). *Right Peripheral Fragments*. Amsterdam: John Benjamins.
- Fernández-Sánchez, Javier, Dennis Ott (2020). "Dislocations. Part I and II". *Language and Linguistics Compass* 14: 1-40.
- Farudi, Annahita (2013). *Gapping in Farsi: A Crosslinguistic Investigation*. PhD dissertation, University of Massachusetts at Amherst.
- Gallego, Ángel J. (2009). "Ellipsis by phase". Paper presented at the XIX Colloquium on Generative Grammar, EHU-UPV.
- Haegeman, Liliane (2006). "Conditionals, factives and the left periphery". *Lingua* 116: 1651-1669.
- Hankamer, Jorge (1979). *Deletion in Coordinate Structures*. New York: Garland Pub.
- Hartmann, Katherine (2000). *Right Node Raising and Gapping: Interface Conditions on Prosodic Deletion*. Amsterdam: John Benjamins.
- Hooper, Joan and Sandra Thompson (1973). "On the applicability of root transformations". *Linguistic Inquiry* 4: 465-497.

- Jayaseelan, Karattuparambil (1990). "Incomplete VP deletion and gapping". *Linguistic Analysis* 20: 39-75.
- Johnson, Kyle (2009). "Gapping is not (VP-)ellipsis". *Linguistic Inquiry* 40: 289-328.
- Jung, Wonsuk (2016). *The Non-unity of Gapping*. PhD dissertation. UPV/EHU.
- Kuno, Susumu (1976). "Gapping: A functional analysis". *Linguistic Inquiry* 7: 300-318.
- Lasnik, Howard (2014). "Multiple sluicing in English?". *Syntax* 17: 1-20.
- Lin, Vivian (2002). *Coordination and Sharing at the Interfaces*. PhD dissertation, Massachusetts Institute of Technology.
- Merchant, Jason (2004). "Fragments and ellipsis". *Linguistics and Philosophy* 27: 661-738.
- Neijt, Anneke (1979). *Gapping: A Contribution to Sentence Grammar*. Dordrecht: Foris.
- Ortega-Santos, Iván, Masaya Yoshida, Chizuru Nakao (2014). "On ellipsis structures involving a wh-remnant and a non-wh-remnant simultaneously". *Lingua* 138: 55-85.
- Reich, Ingo (2006). "Toward a uniform analysis of short answers and gapping". In Denis Pascal (ed.) *Proceedings of the 2004 Texas Linguistics Society Conference*. Somerville, MA: Cascadilla Proceedings Project, pp. 69-78.
- Repp, Sophie (2009). *Negation in Gapping*. Oxford: Oxford University Press.
- Sag, Ivan (1976). *Deletion and Logical Form*. PhD dissertation, Massachusetts Institute of Technology.
- Schneider, Stefan (2007). *Reduced Parenthetical Clauses as Mitigators*. Amsterdam: John Benjamins.
- Toosarvandani, Maziar (2016). "Embedding the antecedent in gapping: low coordination and the role of parallelism". *Linguistic Inquiry* 47: 381-390.
- Vikner, Sten (1995). *Verb Movement and Expletive Subjects in Germanic Languages*. Oxford: Oxford University Press.
- Weir, Andrew (2014). *Fragments and Clausal Ellipsis*. PhD dissertation, University of Massachusetts at Amherst.
- Wurmbrand, Susi (2017). "Stripping and topless complements". *Linguistic Inquiry* 48:341-366.
- Zhang, Niina (2009). *The syntax of coordination*. Cambridge: Cambridge University Press.

---

## **Acknowledgements**

Parts of this paper were originally presented in the Ellipsis Across Borders conference in Sarajevo in 2016, the 37<sup>th</sup> Tabu-DAG at the University of Groningen, as well as a seminar at the Universitat Autònoma de Barcelona in 2017. I am grateful to the audiences of these venues, as well as two anonymous reviewers who challenged and improved my first draft.

Javier Fernández-Sánchez  
ORCID iD: 0000-0002-6060-2332  
Uniwersytet Gdański  
Instytut Anglistyki i Amerykanistyki  
Wita Stwosza 51  
80-308 Gdańsk  
Poland  
javier.fernandez-sanchez@ug.edu.pl





**How language shapes interpersonal distance:  
An analysis of pronominal forms of address  
in Spanish, Polish and Italian**

MAREK BARAN  
EWA URBANIAK

*Received 15.05.2021,  
received in revised form 01.11.2021,  
accepted 02.11.2021.*

**Abstract**

The aim of the present article is to compare the formal and functional aspects of pronominal forms of address in three languages: Spanish, Polish and Italian. The classic typology of the category analysed divides it in two groups: the T-forms applied in the conversations between the participants of symmetrical relations and the V-forms considered reverential and asymmetrical. The present study demonstrates and analyses the pronominal systems in two Romance languages, Spanish and Italian, and a Slavic language, Polish. We classify the pronouns according to the confidentiality/distance parameter, showing the similarities and differences between the formal characteristics, as well as the socio-cultural factors that determine the election of determined pronominal form of address

**Keywords**

pronouns, forms of address, Spanish, Italian, Polish

## **Językowe konfigurowanie interpersonalnego dystansu: analiza pronominalnych form adresatywnych w języku hiszpańskim, polskim i włoskim**

### **Abstrakt**

Celem niniejszego artykułu jest porównanie formalnych i funkcjonalnych aspektów pronominalnych form adresatywnych w trzech językach: hiszpańskim, polskim i włoskim. Klasyczna typologia analizowanej kategorii dzieli ją na dwie grupy: formy typu T stosowane w rozmowach pomiędzy uczestnikami o relacjach symetrycznych oraz formy typu V uznawane za wyraz szacunku i asymetryczności. Niniejsze badanie przedstawia i analizuje systemy pronominalne w dwóch językach romańskich, hiszpańskim i włoskim, oraz jednym słowiańskim, języku polskim. Klasyfikujemy zaimki według parametru poufność/dystans, ukazując podobieństwa i różnice pomiędzy cechami formalnymi, jak i czynnikami społeczno-kulturowymi, które warunkują wybór określonej pronominalnej formy adresatywnej.

### **Słowa kluczowe**

zaimki, formy adresatywne, język hiszpański, język włoski, język polski

### **1. Introduction**

In the course of interaction, participants establish certain relations that are usually reflected by the linguistic mechanisms they apply. One of these interpersonal strategies is the use of the forms of address based on a range of socio-cultural factors that demonstrate the position of each interlocutor in a certain society or within the interaction. In the present article we propose an analysis of the pronominal forms of address used in Spanish, Polish and Italian. The aim of the study is to show what are the formal and functional aspects of the forms of address in these languages. What is more, we propose a socio-pragmatic interpretation of the systems and the changes that they suffer based on Politeness Theory (especially the proposal

of Diana Bravo 2003, 2004) and the concept of dominant interactive style (according to the model proposed and developed by Kerbrat-Orecchioni 1996 and Baran 2010). By contrasting three languages we intend to demonstrate that in every society the pronominal system indicates different social order and socially recognized values. Furthermore, the study shows that typological kinship does not necessarily indicate the similarity between the forms of address systems. As we will see, the Italian and Polish schemes seem far more alike in comparison with the Spanish one.

By applying the model of dominant interactive style, we are referring to the phenomena on which Kerbrat-Orecchioni (1994 2005) bases her theory of *communicative ethos*.<sup>1</sup> Both conversational norms or discursive and interactive mechanisms seem to create a relational network that corresponds to a certain sociopragmatic logic (the term *interpersonal territory* has to be analysed taking into consideration a range of social and pragmatic elements that determine each other). The concept of *communicative ethos* is delimited by the following typological criteria (Baran 2012b: 10-18): 1) quantitative and qualitative weight of a word (verbosity), 2) type of interpersonal relations, 3) linguistic politeness concept, 4) identity concept, 5) level of ritualisation, and 6) emotivity (interpreted as determined interactive strategic actions, and not an individual exteriorization of affective states).

Comparative studies of the speech acts in different linguistic communities or analyses of the linguistic politeness strategies applied in distinct speech communities seem to demonstrate that cultural systems do not differ when it comes to the type of

---

<sup>1</sup> The concept derived from the notion of interpersonal rhetoric that appears, among others, in the works of Leech (1983). It considers a number of interactional strategies that characterize the interactive style of members of a certain speech community. At the same time, those strategies reflect the social and cultural values typical for a certain community. The term *communicative style*, a part of interactional rhetoric, can be interpreted as a set of linguistic actions applied in the course of social communication that characterize a certain linguistic, ethnic or cultural community.

values that are manifested on the level of creating the social universe. What does differ in every culture is the scope of the influence of the values exposed and a specific (frequently hierarchic) relation between the elements. Hernández Sacristán (2003: 39-44) describes the dynamics of the interpersonal relations, constituting a part of a general social dynamics, through the following dimensions:

- a) harmony vs. rivalry,<sup>2</sup>
- b) solidarity vs. the non imposing principle,<sup>3</sup>
- c) authenticity vs. ceremoniality,<sup>4</sup>
- d) affectivity vs. restraint principle,<sup>5</sup>
- e) liberty vs. obligatority.<sup>6</sup>

Taking under consideration the premises presented, we sustain that the forms of address can be considered a very sensitive parameter that demonstrates the dynamics of such relations. Consequently, a contrastive, typological-functional analysis of the forms of address pronouns that we undertake in this article can reflect the real impact of the principles of: harmony, solidarity, authenticity or affectivity (and their contrary elements). In this respect, the aim of the present study is:

---

<sup>2</sup> Although only on the strategical dimension, the conversational expression of agreement and disagreement can be influenced by different sociocultural parameters: the level of assumed verbally manifested unanimity is related to a minor or major representation of the principle of harmony.

<sup>3</sup> In some cases, the universal principle of communicative cooperation has to include, for example, the scope of individual autonomy (the concept of interactive territory is conceptualised heterogeneously).

<sup>4</sup> The level of ritualisation of the verbal exchange is not equal in every speech community.

<sup>5</sup> Emotivity of the communicative acts undergoes some cultural scripts. Verbal and paraverbal codes and some proxemic signs demonstrate the heterogeneity of the ways of exteriorizing emotions and feelings.

<sup>6</sup> Existence of certain social norms, constituting some schemes of interactional behaviour, does not exclude the exceptions, that can be very differently evaluated.

- to mark the sociopragmatic conventions that determine the conversational-interactional actualization of the pronominal forms of address in the analysed speech communities;<sup>7</sup>
- to determine to what degree the communities classified as representants of the positive politeness model (what concerns the three communities examined) have developed not always corresponding mechanisms, by implication we investigate the social perception of the distance/closeness parameter.

The present study is not a corpus study, although we do evoke some examples from corpora and other investigators' works.

## 2. Forms of address

The basis for numerous studies on the pronominal forms of address is Brown and Gilman's theory (1960) which introduces the conceptual opposition of *power* and *solidarity* conceived from a psychosocial perspective. The concept of *power* reflects the control that certain people take (or may take) over others in a particular interactive situation. In that case, the interlocutors tend to apply the asymmetrical pronouns – the part that is in control uses a more confidential form (like *ty* in Polish or *tú* in Spanish), while the person that is controlled uses some reverential forms (like *pan/pani* or *usted*). The category of *power* is provoked by a range of socio-cultural factors such as position in a social or economic hierarchy, age or sex. The *solidarity* parameter reflects the situation in which both of the speakers occupy the same social position and maintain certain kind of relation. In that case, they apply the symmetric formulas like *ty* in Polish or *tú* in Spanish.

Throughout the years societies have changed so the classical dichotomy should be enriched in the third, special type of *power* relation – the situation in which one of the speakers occupies a higher position in the social hierarchy than the

---

<sup>7</sup> We use the term *speech community* on purpose, as, when it comes to Spanish, our analysis is limited to the European variety – Peninsular Spanish.

other, but, according to the generally accepted rules, they both apply the V forms (like *usted* or *pan/pani*). The typical asymmetrical relation of power seems to have disappeared in the majority of social contexts, although it is maintained in some very specific situations (for example, while talking to children).

Brown and Gilman's theory constitutes a basis for the research of forms of address, though we consider it should be revised and in some aspects modified. Firstly, as we have already depicted, we consider the categories of *power* and *solidarity* as dynamic and constantly changing due to the social changes and the social perception of hierarchic positioning. This means that the research on the use of pronominal forms should be regularly completed and developed. What is more, the analysed parameters are interactionally dependent – not only is the social position that affects the possible relation, but there are also some interactional parameters that should be taken under consideration while analysing concrete pronominal uses, such as time and place, the presence of other participants or the general purpose of the interaction, among others. As an example of the interactional dependence we can mention the academic situation in which two professors who occupy the same hierarchy position and apply the solidarity T-forms, during a faculty reunion, switch to the V-forms as it is commonly accepted verbal behaviour.

We should also outline that the *power* and *solidarity* parameters are strictly related to a concrete socio-cultural environment. That means that even though the T/V dichotomy appears in two language systems, it is not always applied under the same conditions. We believe that societies can be divided into two types: in which prevail the hierarchical order and the ones with the dominating solidarity social structure. Each of these tendencies is reflected by a certain distribution of T/V pronouns.

The question of pronominal forms of address is frequently associated with politeness theory, introduced by Brown and Levinson (1978) and developed by a great number of special-

ists (Kerbrat-Orecchioni 1996, Bravo 2003, 2004, Hernández Flores 2002, 2004, Kita 2005, Albelda and Barros 2013, among others) assumes that the interaction is an interpersonal play in which every participant is trying to prevent their and their interlocutor's *face*, constantly threatened by the *Face Threatening Acts*, by introducing some specific verbal expressions. There are two types of verbal politeness: positive (which prevents the speakers from being excluded from a certain group) and negative (which prevents them from being imposed by other participants of an interaction). According to that theory, the T-forms could be considered as positive politeness phenomena (as they provoke a feeling of solidarity and equality between the speakers), while the V-forms represent the negative politeness (they help to keep the distance between the interlocutors). Nevertheless, our intention is to show that T/V opposition does not always reflect the positive/negative politeness dichotomy. Recent politeness theory studies show that politeness rules vary according to the situation, context and society – which means that, under certain circumstances, also the V-forms can be considered as solidarity pronouns as they reflect that the speakers share the same values and belong to the same group.

In the following sections we will briefly discuss the formal and socio-pragmatic aspects of the pronominal forms of address in three languages: Spanish, Italian and Polish. As we observe, they share a number of similarities, both in the organization and the interactional meaning. Nevertheless, we will also demonstrate that every society builds its own particular hierarchy system which is reflected in the use of particular pronouns during everyday conversations.

### **3. Spanish, Italian and Polish: systems of pronominal forms**

In the present work we will apply the *confidence/distance* parameters that we believe demonstrate the actual state of art.

The confidence pronouns reflect a close relation between the speakers, while the distance pronouns mark a less intimate relation. The distance between the speakers, as we intent to demonstrate, depends not only on many socio-pragmatic variables, like the position in hierarchy, the familiar bounds or the age, but it is also created in the course of interaction. What is more, the *confidence/distance* opposition reflects the inner structure of a certain community. That means that some communities are more eager to apply the confidence or distance forms, because they reflect their dominant interactive style.

The Spanish system is known for the diversity of pronouns systems depending on the diatopic factors (see, among others, Almeida and Mendoza, 1992; Betolotti, 2015; Sampedro, 2021). Concerning the geographical extent of the Spanish dominium, it comes as no surprise that it presents a variety of subsystems particular for a determinate community. The dominant system in many parts of Spain and included in the “standard” version of Spanish divides the pronouns the following way:

**Table 1**  
Spanish I

	SINGULAR	PLURAL
Confidence	Tú	Vosotros/as
Distance	Usted	Ustedes

This is the only system in which the confidence/formality parameter is reflected in the plural pronouns, as in the rest of the systems those notions seem neutralized (there is only one form *ustedes* used in both confidential and formal situations). Apart from the “standard” system, frequently taught during Spanish classes to the non-native speakers, there are several others schemes, all collected in the work of Fontanella de Weinberg (1999: 1401-1408).



**Table 2**  
Spanish II

	SINGULAR	PLURAL
Confidence	Tú	Ustedes
Distance	Usted	Ustedes

The second system is characteristic for some of the parts of the Iberic Peninsula, such as western Andalucía, some parts of the regions of Córdoba, Jaen and Granada, in Canary Islands, Mexico, Peru, a vast part of Colombia, Venezuela and the Antilles.

**Table 3**  
Spanish IIIa

	SINGULAR	PLURAL
Confidence	Vos ~ Tú	Ustedes
Distance	Usted	Ustedes

**Table 4**  
Spanish IIIb

	SINGULAR	PLURAL
Intimicy	Vos	Ustedes
Confidence	Tú	Ustedes
Distance	Usted	Ustedes

The third systems can be divided into two types. In both of them the pronoun *vos* is present, although indicating slightly different notions. According to Fontanella de Wienberg (1999: 1404-1405) the IIIa type can be encountered in the area of Chile, Bolivia, the south of Peru, some parts of Colombia, western Venezuela, the region between Panamá y Costa Rica and a Mexican state Chiapas. The type IIIb, on the other hand, is characteristic for Uruguay. As we can see, the Uruguayan system consists of three elements, therefore three levels of confidentiality. Like in other cases, the pronoun *usted* reflects the higher level of formality. When it comes to the opposition *vos/tú*, the first one is used between the speakers of a high

level of intimacy, while *tú* demonstrates more confident though not that intimate relations.

**Table 5**  
Spanish IV

	SINGULAR	PLURAL
Confidence	Vos	Ustedes
Distance	Usted	Ustedes

The last system is often used in Argentina, Costa Rica, Nicaragua, Guatemala and Paraguay. This particular type does not contain the form *tú* which is substituted by *vos*.

When it comes to the Italian system, it consists of the following elements:

**Table 6**  
Italian

	SINGULAR	PLURAL
Confidence	Tu	Voi
Distance	Voi/Lei/Ella	Voi/Loro

Although it does not present such diversity as the Spanish system, some of the forms are problematic. Firstly, the form *lei* presents ambiguity as it manifests two meanings – it is not only an addressative pronoun that indicates formality, but can also indicate third person, singular, feminine as in the sentence *lei è stanca* – ‘she is tired’. As an addressative pronoun it can be applied both to a male or female interlocutor: *Signora Rossi, lei lavora a Roma?*/*Signor Rossi, lei lavora a Roma?* (‘Signora Rossi, you work in Rome?/Signor Rossi, you work in Rome?’).

Among the formal singular forms, it is *lei* which is considered the most typical realization of *italiano standard*, while the *ella* form is highly formal, used mainly in some administrative texts. According to Bresin (2019), *voi* can be interpreted as an archaism or a French borrowing. Nevertheless, some linguists like Niculescu (1974: 26) consider the *voi* form as a regional-

ism applied in some southern parts of Italy. In the regions where both of the forms *voi* and *lei* are maintained, *lei* is considered as more formal, applied in some situations that require the highest level of reverence. A parallel situation has been observed in the *voi/loro* opposition as plural forms of address. As acknowledged by many (Inglese 2002, Sobrero 2011, Formentelli and Hajek 2015), the pronoun *loro* is used exclusively in some highly formal situations, while *voi* is the most common form addressed as a reverence indicator to multiple interlocutors.

Table 7 presents the Polish pronoun system.

**Table 7**

Polish

	SINGULAR	PLURAL
Confidence	Ty	Wy
Distance	Pan/pani	Państwo

As we can observe, Polish pronominal system does not vary diatopically. On the other hand, the forms *pan/pani/państwo* present some formal ambiguity (see Wierzbicka 2016). Most grammars consider *pan/pani/państwo* not as pronouns, but as nouns: for example, in many contexts, these forms can be translated as *Sir/Madame* (for example, while accompanying names or surnames in phrases like *pani Kasia*, *państwo Kowalscy*, and they can also be used as equivalents of *a woman* or *a man* in sentences like *Rozmawiałem z tamtym panem* – ‘I have talked to that man’). Nevertheless, many investigators, Łaziński (2006:15-17) among them, considers *pan/pani/państwo*, as pronouns that express verbal politeness (such is the case in the sentences like *Czego się pani napije?* – ‘What will you (formal form) drink?’).

## 4. Analysis

As mentioned above, the confidentiality/distance parameter is present in all of the languages analysed. What significantly differs is the situational context in which they appear. In these three languages the distance forms are used in highly formal situations in which it is of a great value to mark the distance between the interlocutors. We refer to most of the public acts in which interact the representants of some respectful institutions, such as ministers, chairmen, politicians, etc. Nevertheless, in numerous situations, the confidentiality/distance parameter is applied differently, according to a certain sociocultural pattern. In this section we present a selection of contexts in which every community seems to present some peculiarities. Although the examples presented do not encompass all of the possible differences, we believe that they reveal some characteristics of a dominant interactive style of the communities described.

### 4.1. Spanish *tuteo*

During the last decades, a growing tendency of *tuteo* (the use of T-forms) is observed, especially in Peninsular Spanish<sup>8</sup> (Wolarska 2004, Granvik 2007). This means that in all types of conversations, also with the speakers that occupy a high position in the social hierarchy, it is totally acceptable to apply the T-forms. Granvik (2007: 238) demonstrates that the distance forms are used only while speaking to the elderly. This indicates that the factor of social hierarchy seems to lose its strength when it comes to the Spanish speaking society.

At this point, Polish and Italian communities seem to greatly differ from the Spanish speakers. In both communities, while speaking with the representants of higher social positions the V-forms are the most appropriate ones. Among the contex-

---

<sup>8</sup> In other regions the V-forms are still maintained more frequently than in Peninsular Spanish.

tual situations in which the Spanish systems significantly differs from the Italian and Polish ones there are the conversations between the teachers and students at school and Universities. In Italian and in Polish it is obligatory for students to apply the V-forms, despite other socio-pragmatic factors (like age – even though the teacher is very young, the students are obliged to use the *lei* or *pan/pani* pronouns). What is more, in order to emphasize the hierarchy position, the Polish pronominal system is often complemented with the specific titles usually related to one's office. Such a phenomena is called *tytułomania* - 'obsessions of titles', because the frequency of use of the professional titles seems higher than in other speech communities (Bogusławski 1996: 84-85, Łaziński 2006: 137-138, Huszcza 2005, Baran 2012a). In that respect, the Polish system seems highly asymmetrical – one participant of the conversation marks the higher position of their interlocutor, while the other one uses only the pronominal forms. The phenomena described is especially characteristic for the academic settings – the students are obliged to use not only the pronominal reverence form, but also the academic titles. However, as Łaziński (2006: 76-78) mentions, the use of academic titles seems less frequent than in the past, by complementing the pronominal system with the nominal structures a higher level of distance is accomplished.

In Spanish, on the other hand, the scholastic context does not require the use of *usted* – on the contrary, the pronoun *tú* is applied both by the students and the teachers (Sampedro 2021). In higher education institutions, like Universities, V-forms are used, if the professor with whom the interaction is undertaken is highly respectful or much older than the students. Nevertheless, in a great majority of University situations (during classes, tutorials, etc.) the student-professor interactions are based on the symmetrical T-forms.<sup>9</sup>

---

<sup>9</sup> The T-forms applied in scholar systems are well demonstrated in literature, movies and series and constitute a difficult challenge for the Polish translators. For example, in one of the Netflix series "El desorden que dejas".

The other great difference between the use of confidence/distance forms in the analysed language systems regards situations in which the interlocutors are strangers, do not know each other. In Polish and Italian communities we observe the tendency to apply the V-forms during different interactions usually in public surroundings (in restaurants, bars, shops, supermarkets, etc.). The socio-pragmatic rules seem to change when the interlocutors are both relatively young – more and more often we can witness the T-forms in conversations between strangers that are of the same age. Nevertheless, still the most frequent and “safe” form that will certainly not offend anyone are the V-forms. Grybosiova (1998) sustains that such changes are based on the extralinguistic, sociocultural premises and should be studied taking under consideration new patterns followed by the societies.<sup>10</sup>

In Spanish, on the other hand, the factor of knowing or not knowing the interlocutor does not influence to such a degree the election of the form. It is generally accepted to use the T-form while speaking in public spaces, except from the situations in which the interlocutors are willing to mark the distance. Roselló (2018: 256), who analyses the forms of address

---

the main character is a teacher, so a great part of the series consists of teacher-students interactions. The Polish translators have maintained the T-forms, although it does not necessarily reflect the Polish communicative style. On the other hand, it seems that, when possible, they introduced some impersonal forms. For example, when one student asks the teacher “¿Me puedes contestar una pregunta?” (“Can you give me an answer to one question?”) it is translated into “Mogę o coś zapytać?” (“Can I have a question?”) which avoids using any forms of address. We believe that such changes reflect the difficulties encountered while translating the Spanish conversations in this particular context.

<sup>10</sup> For Grybosiova, the change from T-forms to V-forms is a result of a fascination of Polish society with the English culture and language. She sustains that the general patterns followed by the society are “foreign=good” and “new=good”. Although the influence of English and American cultures is undeniable, we believe that the societies, cultures, languages and interpersonal relations change, because the change is one of their internal characteristics. There might be some influence of the expansion of English in the world, although, we suspect that it cannot be considered a dominant factor, as, despite the impact of foreign cultures, the interactive style is a dynamic concept that constantly evolves during everyday conversations.

basing on the oral corpus PRESEEA, affirms that the data collected confirm the unmarkedness of the T-form respect to the V-form. As he states, “*tú* es la forma no marcada, la más usual al dirigirse a la otra persona” (“*tú* is the unmarked form, the most usual when speaking directly to other person”). Sometimes the unknown interlocutors apply the V-forms, for example, if the age gap between the interlocutors seems significant or if at least one of the participant’s intention is to underline the distance. Such is the case of the places that wish to be considered as luxurious, like some restaurants or shops with expensive products. In those establishments the waiters or sales assistants are used to apply the V-forms, even while speaking with young interlocutors. Nevertheless, in such situations the pronouns *usted/ustedes* seem to marked, introducing some extra interactive and social meanings.

#### **4.2. Plural forms**

The pronominal forms of address system in three languages analysed present a certain peculiarity – a diminishing confidence/distance dichotomy in the pronouns directed to a multiple recipient. In the Spanish standard system, we can observe the opposition between *vosotros* that include the confidence notion and *ustedes* which marks a certain distance. Nevertheless, as we can see in the tables above, the rest of the systems does not include such a dichotomy applying only one form in both context – *ustedes*. It seems interesting that the form chosen is the reverence one. Nevertheless, we believe that regardless of the form, the simplification of the system reinforces the solidarity relations between the speakers – it indicates that, while talking to the plural recipient, all of the variables that are taken under consideration do not apply. Probably, some sociocultural factors are reflected by other linguistic or discourse mechanisms (like lexical units or the topics that are raised in the course of interaction). Nevertheless, the pro-

nominal forms of address in a majority of Spanish speaking areas do not contain the confidence/distance opposition.

The Italian system formally marks the confidentiality/distance opposition by the pronouns *voi/loro*. Nevertheless, the *loro* form is nowadays considered as highly formal and it is applied only in the context of the highest level of respect (Scaglia 2003, Sobrero 2011). It means that in situations in which the use of distant *lei* is required, while speaking with a plural interlocutor, the distant form is substituted with a confidential *voi*

When it comes to the Polish system, the standard sociocultural norms indicate that the pronouns addressed to a plural recipient should formally mark confidence/distance dichotomy. Nevertheless, the everyday uses seem to reflect a constant change in the perception of the adequate forms. Firstly, we can observe an intermediate phenomena which consists in using the *państwo* pronoun with a verb in second person plural, like in the sentence *Zadzwońcie państwo o siódmej*. It can be applied in very formal contexts, like in the following example:

Zwróćcie państwo uwagę, że jeżeli taką ustawę przyjmiemy, jej realizacja może być dla społeczeństwa wielce kontrowersyjna, zwłaszcza w kontekście wojny z Iranem.<sup>11</sup>

The fragment above constitutes a part of speech during the Parliament session. In such a highly formal situation, the speaker uses the form “zwróćcie uwagę” (‘pay attention’) using the verb in second person plural form and a V-pronoun. Such a peculiar mechanism in which the verb does not conjugate according to the pronoun applied is recently quite often in some semi-formal situations (in which the V-form seems too distant and the T-form too confidential). It also occurs if the speaker for some reasons wishes to reduce the interpersonal

---

<sup>11</sup> NKJP, Sprawozdanie z 8. posiedzenia Senatu RP część 2, wersja robocza, 4. Kadencja, <http://www.nkjp.uni.lodz.pl> (17/11/2021).



distance maintained with other participants of the conversation.

It is also possible to apply the pronoun *wy* in situations in which it is the V-form that is considered the most appropriate due to the socio-cultural norms. For example, it can be observed that while talking to strangers, the Polish speaking participants apply the pronoun *wy* as a sign of reducing the distance and making the conversation less formal. It can be observed in the fragment below:

- *Tak. By w 1/16 nie trafić na słynnego Szweda Ljungberga. Co się opłaciło i skończyło srebrnym medalem.*

\* *Co robicie, by zmienić przepisy, który prowadzą do prostytucji sportu?*

- *Piszemy, posyłamy projekty a działacze FILA milczą.*<sup>12</sup>

This is a fragment of an interview with sports coach Józef Tracz published in *Gazeta Wroclawska*. The journalist uses the T-form *wy* (“co robicie” – ‘what are you doing’) and not the V-form *państwo*, probably in order to reduce the distance and make the conversation more direct.

This brief analysis of the forms used in three languages shows us that the plural forms of address present the tendency to reduce the opposition confidentiality/distance by limiting the use of the pronouns that reflect the distance. Every language does it to a different degree and with different strategies. The Italian system presents the highest level of reduction, as generally it is the T-form that is used in everyday language, while the V-form is applied in very strict, highly formal situations. The Spanish standard system does include the *vosotros/ustedes* dichotomy, but many dialectal variations limit it to one form *ustedes*. The Polish system seems less susceptible to change, as generally it is the form *Państwo* that is considered the most adequate while speaking to the interlocutors

---

<sup>12</sup> NKJP, Dlaczego zapaśnik walczy, by przegrać? Odpowiada trener Józef Tracz, <http://www.nkjp.uni.lodz.pl> (17/11/2021).

with whom the distance should be marked. On the other hand, we observe a spreading tendency to apply some specific mechanisms (like the use of verb in the second person plural) in order to underline, perhaps not confidentiality, but at least a reduction of interpersonal distance between the speakers.

### **4.3. Switching from V-forms into T-forms**

As it has been observed, the Spanish pronominal forms of address system differs greatly from the Polish and Italian ones. Among the differences, there is one characteristic that should be mentioned – a way in which the speakers change the forms of address reducing the distance between them. As it has been observed in the Spanish speaking communities, in Spanish the process of modifying the forms of address applied occurs in the course of interaction – the speakers fluently change from *usted* to *tú* pronouns (Blas Arroyo 1994: 43-409, 2005: 318-319, Baran 2012a: 39-41). During one interaction the participant may start with the forms of distance, then pass to the asymmetrical forms (one participant applies the V-form and the other one the T-form) and then they switch to the symmetrical use of confidential forms. In that case, a change from the distance to the confidential forms acts as a contextualization indicator – a modification in the perception of the interaction context is reflected by the change of the forms applied.

On the other hand, in Polish and Italian systems the change of pronominal forms of address is considered a specific ritual accompanied by specific verbal expressions. In both communities, it is the person of a higher social status that is considered the most adequate to initiate the ritual (Benigni and Bates 1977: 159, Renzi *et al.* 2001: 373, Marcjanik 2009: 36-38). What is more, there are specific formulas that introduce the ritual: like *Przejdźmy na ty* in Polish or *Diamoci del tu?* in Italian (they can be translated as ‘Shall we use the T-form?’). In

contrast with the Spanish convention,<sup>13</sup> in Polish and Italian communities a change of form of address constitutes a deliberate decision based on a specific ritual.

#### **4.4. Socio-pragmatic interpretation**

It is a generally observed tendency of the western societies to flatten the social hierarchy ladder what can be easily detected in the linguistic forms and strategies applied. The comparative analysis shows us that within the European societies, each one presents their own peculiarities. The Spanish society seems to represent the equality model in which the solidarity pronouns are preferred. The Polish and Italian systems, on the other hand, still maintain the importance of hierarchic position, although we can observe some mechanisms and strategies in the contemporary use that demonstrate some exemptions from a traditional hierarchic model.

It is curious though that the Italian system is more similar to the Polish than to the Spanish one taking under consideration the typological kinship of the languages. It seems that it is not the typological background that has the major impact on the organization of pronoun forms of address, but the socio-cultural values that are appreciated in a certain speech community.

Although the Spanish community represents the equality model, while Italian and Polish systems opt for the hierarchical one, we strongly believe that it does not reflect the psychological characteristic of the societies in question. Every community applies a certain interactive style which is developed throughout the years. The Spanish community has uncon-

---

<sup>13</sup> As demonstrated by Blas Arroyo (1995: 234-235), in Spanish sometimes the selection of a certain form depends on the individual decisions of the speakers. The author cites an example of a conversation during university class between professor and two students: one was applying the T-form and the other one the V-form while referring to the professor. There is no special ritual that introduces the T-form, but a dynamic decision based on the interpersonal perspective adopted by the interlocutors.

sciously chosen the confidentiality model in which the speakers use the mechanisms and strategies that reduce the distance between the speakers. The Polish and Italian systems, on the other hand, are based on the distance interactive style in which the confidential forms are reserved for generally close relations between the speakers. It does not mean that the Spanish society could be evaluated as more nice or polite – the reduction of distance should be simply interpreted as a dominating interactive style.

The politeness theory suggests that in every society there is a set of verbal behaviour rules. As Kerbrat-Orecchioni (2004) assumes, the existence of politeness forms is common for every society, but the rules may differ according to the cultural aspects. In order to avoid the *Face Threatening Acts* and putting in danger the interactional success of the conversation, the speakers subordinate to the rules valid in a particular context. If we compare the same situation in two different societies, possibly the politeness strategies applied will be significantly different. For example, we can imagine a typical situation in which one person asks a stranger on the street how to get to a certain place. According to the Italian and Polish politeness norms, the most adequate strategy would be to use the V-forms (*lei* in Italian or *pan/pani* in Polish). Spanish norm, on the other hand, accepts the T-form represented by the pronoun *tú*. It reflects that the dominant interactive style in Polish and Italian maintains certain distance between the speakers that does not know each other, while the Spanish system admits more confidential forms. Although it does not reflect the psychological characteristics of the societies in question, it certainly demonstrates a general interactive pattern.

The use of the V/T-forms according to the politeness rules established in a particular society can be considered a mechanism of a socio-pragmatic concept of *social image of affiliation* proposed by Bravo (2003). By applying the forms (both of confidentiality and distance) due to the socio-pragmatic norms, the speaker demonstrates their wish to belong to a certain co-

community and to be seen as a part of it. Both using a distant form in situations that require the confidential ones and applying the confidential forms that are considered inappropriate in a certain context can be interpreted as the threat on the social image of affiliation of the speaker. What differs one community from the other are the socio-cultural factors that determine on which values the generally accepted politeness rules are to be based.

## References

- Albelda Marco, Marta, Jesús Barros García (2013). *La cortesía en la comunicación*. Madrid: Arco Libros.
- Almeida Suárez, Manuel, Juana Mendoza (1992). "Formas pronominales de tratamiento en español actual". In: Jeanine Stolidi (ed.). *Recherches en linguistique hispanique: Actes du colloque d'Aix-en-Provence, 20 et 21 Mars 1992*. Université de Provence, 167-178.
- Baran, Marek (2010). *Emotividad y convención sociopragmática: Una contribución al estudio del ethos comunicativo de la comunidad hispanohablante peninsular*. Łódź: Wydawnictwo Uniwersytetu Łódzkiego.
- Baran, Marek (2012a). "Formas nominales y pronominales de tratamiento: perspectivas de estudios contrastivos en el ámbito sociopragmático". *Studia Iberystyczne* 11: 27-56.
- Baran, Marek (2012b). "Teoria etosu komunikacyjnego a opis socjopragmatyczny". In: Jarosław Płuciennik, Piotr Stalmaszczyk (eds.). *Metody badań i poszukiwania*. Łódź: Primum Verbum, 51-69.
- Benigni, Laura, Elizabeth Bates (1977). "Interazione sociale e linguaggio: Analisi pragmatica dei pronomi allocutivi italiani". In: Raffaele Simone, Giulianella Ruggiero (eds.). *Aspetti socio-linguistici dell'Italia contemporanea*. Roma: Bulzoni, 141-65.
- Bertolotti, Virgini (2015). *A mí de vos no me trata ni usted ni nadie: Sistema e historia de las formas de tratamiento en la lengua española en América*. México: Universidad Nacional Autónoma de México.

- Blas Arroyo, José Luis (1994). "De nuevo sobre el poder y la solidaridad: Apuntes para un análisis interaccional de la alternancia Tú/Usted". *Nueva Revista de Filología Hispánica* 42: 385-414.
- Blas Arroyo, José Luis (1995). "Un ejercicio de sociolingüística interaccional: El caso de los pronombres en el tratamiento del español actual". *Verba* 22: 229-252.
- Blas Arroyo, José Luis. (2005). *Sociolingüística del español: Desarrollos y perspectivas en el estudio de la lengua española en contexto social*. Madrid: Cátedra.
- Bogusławski, Andrzej (1996). "Deutsch und Polnisch als Beispiele eines «egalitären» und eines «antiegaltären» Anredesystems". In: Wolfgang Gladrow, Sonja Heyl (eds.). *Slawische und deutsche Sprachwelt. Typologische Spezifika der slawischen Sprachen im Vergleich mit dem Deutschen*. Frankfurt am Main: Berlin – Bern – Bruxelles – New York – Oxford – Wien: Peter Lang, 78-86.
- Bravo, Diana (2003). "Actividades de cortesía, imagen social y contextos socioculturales: Una introducción". In: Diana Bravo (ed.). *Actas del Primer Coloquio del Programa EDICE: Perspectiva no etnocentrista de la cortesía: Identidad sociocultural de las comunidades hispanohablantes*. Stockholm: Stockholms Universitet, 98-107.
- Bravo, Diana (2004). "Tensión entre universalidad y relatividad en las teorías de cortesía". In: Diana Bravo, Antonio Briz Gómez (eds.). *Pragmática sociocultural: Estudios sobre el discurso de cortesía e español*. Barcelona: Ariel Lingüística, 15-37.
- Bresin, Agnese (2019). "Perception of address practices in Italian interregional encounters: A case study of restaurant encounters". *Journal of Pragmatics* 143: 185-200.
- Brown, Penelope, Stephen Levinson (1987). *Politeness: Some Universals in Language Usage*. Cambridge: Cambridge University Press.
- Brown, Roger, Albert Gilman (1960). "The pronouns of power and solidarity". In: Thomas A. Sebeok (ed.). *Style in Language*. New York: Wiley, 253-276.
- Fontanella de Weinberg, María Beatriz (1999). "Sistemas pronominales de tratamiento usados en el mundo hispánico". In: Ignacio Bosque, Violeta Demonte (eds.). *Gramática Descriptiva de la Lengua Española*. Madrid: Espasa Calpe, 1400-1425.
- Formentelli, Maicol, John Hajek (2015). "Address in Italian academic interactions: The power of distance and (non)-reciprocity". In: Catrin Norrby, Camilla Wide (eds.). *Address Practise as Social Ac-*

- tion: European Perspectives*. Hampshire: Palgrave Macmillan, 119-140.
- Granvik, Anton (2007). "Formas de tratamiento e interferencia: Estudio sobre el uso de las formas de tratamiento españolas por parte de portugueses nativos residentes en Madrid". *Revista de Filología Románica* 24: 221-250.
- Grybosiowa, Antonina (1998). "Przyczyny zmian w polskim systemie adresatywnym". *Acta Universitatis Lodzianensis: Folia Linguistica* 37: 57-61.
- Hernández Flores, Nieves (2002). *La cortesía en la conversación española de familiares y amigos: La búsqueda del equilibrio entre la imagen del hablante y la imagen del destinatario*. Aalborg: Aalborg Universitet.
- Hernández Flores, Nieves (2004). "La cortesía como la búsqueda del equilibrio de la imagen social". In: Diana Bravo, Antonio Briz Gómez (eds.). *Pragmática sociocultural: Estudios sobre el discurso de cortesía e español*. Barcelona: Ariel Lingüística, 95-108.
- Hernández Sacristán, Carlos (2003). "Interculturalidad, transculturalidad y valores de la acción comunicativa". In: Grupo CRIT. *Claves para la comunicación intercultural: Análisis de interacciones comunicativas con inmigrantes*. Castelló de la Plana: Universitat Jaume I, 37-87.
- Huszczka, Romuald (2005). "Politeness in Poland: From 'titlemania' to grammaticalised honorifics". In: L. Hickey (ed.). *Politeness in Europe*. Clevedon; Buffalo: Multilingual Matters, 218-233.
- Inglese, Mario (2002). "Sintassi e pragmatica del pronome loro". *Italica* 79: 466-478.
- Kerbrat-Orecchioni, Catherine (1996). *La conversation*. Paris: Seuil.
- Kerbrat-Orecchioni, Catherine (2004). "Es universal la cortesía? Tensión entre universalidad y relatividad en las teorías de cortesía". In: Diana Bravo, Antonio Briz Gómez (eds.). *Pragmática sociocultural: estudios sobre el discurso de cortesía e español*. Barcelona: Ariel Lingüística, 39-53.
- Kerbrat-Orecchioni, Catherine (2005). *Le discours en interaction*. Paris: Armand Colin.
- Kita, Małgorzata (2005). *Językowe rytuały grzecznościowe*. Katowice: Wyższa Szkoła Zarządzania Marketingowego i Języków Obcych.
- Leech, Geoffrey (1983). *Principles of Pragmatics*. London: Longman.

- Łaziński, Marek (2006). *O panach i paniach: Polskie rzeczowniki tytułowe i ich asymetria rodzajowo-płciowa*. Warszawa: Wydawnictwo Naukowe PWN.
- Marcjanik, Małgorzata (2009). *Mówimy uprzejmie: Poradnik językowego savoir-viver'u*. Warszawa: PWN.
- Niculescu, Alessandro (1974). *Strutture allocutive pronominali reverenziali in italiano*. Firenze: Olschki.
- Renzi, Lorenzo, Giampaolo Salvi, Anna Cardinaletti (eds.) (2001). *Grande grammatica italiana di consultazione III*. Bologna: Il Mulino.
- Roselló Verdeguer, Jorge (2018). "Las formas de tratamiento en el corpus PRESEEA-Valencia: Un estudio sociolingüístico". *Círculo de Lingüística Aplicada a la Comunicación* 76: 241-260.
- Sampedro Mella, María Carmen (2021). "Las formas de tratamiento en la tradición académica del español". *Anuario de Letras. Lingüística y Filología* 9: 105-132.
- Scaglia, Claudia (2003). "Deissi e cortesia in italiano". *Linguistica e Filologia* 16: 109-145.
- Sobrero, Alberto (2011). "Pragmatica". In: Alberto A. Sobrero (ed.). *Introduzione all'italiano contemporaneo: Le strutture*. Roma – Bari: Laterza, 403-450.
- Wierzbicka, Anna (2016). "Terms of address as keys to culture and society: German *herr* vs. Polish *pan*". *Acta Philologica* 49: 29-44.
- Wolarska, Agata (2004). *Pragmalingwistyczna analiza zwrotów adresatywnych w języku hiszpańskim i polskim: Studium porównawcze*. PhD dissertation, Poznań, Uniwersytet im. Adama Mickiewicza.

Marek Baran  
ORCID iD: 0000-0001-7728-8733  
Wydział Filologiczny  
Uniwersytet Łódzki  
ul. Pomorska 171/173  
90-236 Łódź  
Poland  
marek.baran@uni.lodz.pl



Ewa Urbaniak  
ORCID iD: 0000-0002-5298-8335  
Wydział Filologiczny  
Uniwersytet Łódzki  
ul. Pomorska 171/173  
90-236 Łódź  
Poland  
ewa.urbaniak@uni.lodz.pl



**Uniqueness in languages  
with and without articles:  
Catalan vs Russian**

DARIA SERES  
JOAN BORRÀS-COMES  
M. TERESA ESPINAL

*Received 05.07.2021,  
accepted 01.11.2021.*

**Abstract**

The article compares the interpretation of singular topical nominals in Romance (Catalan) and Slavic (Russian), and its relation to the presence/absence of the article in the overt morphosyntax. The empirical study, presented in this paper, confirmed the theoretical prediction that in Catalan the presence of a definite article conveys uniqueness of the referent, while an indefinite article suggests non-uniqueness. In the absence of articles (in Russian), bare nominals are compatible with both a uniqueness and a non-uniqueness interpretation. The reading of a bare noun phrase is inferred pragmatically, depending on contextual factors and the background knowledge of the interlocutors.

**Keywords**

noun phrase, articles, definiteness, uniqueness

## **Unikatowość w językach z i bez rodzajników: kataloński vs rosyjski**

### **Abstrakt**

Artykuł porównuje interpretację liczby pojedynczej rzeczownika w języku romańskim (katalońskim) i słowiańskim (rosyjskim) oraz jej związek z obecnością/nieobecnością rodzajnika w morfoskładni. Badania empiryczne przedstawione w niniejszym artykule potwierdziły przewidywania teoretyczne, że w języku katalońskim obecność rodzajnika określonego świadczy o unikatowości desygnatu, podczas gdy rodzajnik nieokreślony sugeruje brak unikatowości. W przypadku braku rodzajników (w języku rosyjskim), rzeczowniki są zgodne zarówno z interpretacją unikatową, jak i nieunikatową. Czytanie bezrodzajnikowej frazy nominalnej jest dokonywana pragmatycznie, w zależności od czynników kontekstowych i ogólnej wiedzy rozmówców.

### **Słowa kluczowe**

fraza nominalna, rodzajniki, określoność, unikatowość

### **1. Introduction**

This article focuses on the way the uniqueness of a referent is encoded/decoded in Catalan (a Romance language with articles) and in Russian (a Slavic language without articles). The uniqueness interpretation that a referent of a nominal gets is generally associated with its definiteness, expressed by a definite article in languages with articles. However, when it comes to languages that do not have articles as a lexical category, the readings that nominals may get are not that clear. And the long-standing debate in linguistics on whether the grammatical category of (in)definiteness exists in these languages is brought about. Considering that about half of languages in the world do not have articles (Longobardi 2001; Dryer 2013a, 2013b, among others), the overt marking of definiteness as

uniqueness-encoding does not seem to be crucial for human linguistic interaction (Lyons 1999). Nonetheless, the distinction between a unique (definite) and a non-unique (indefinite) reference is important for human communication and is, thus, expected to be universally present in language (Cummins 1998; Lyons 1999; Brun 2001, among others).

In relation to the debate on the universality of (in)definiteness, this article compares readings of Russian singular bare nominals that may be interpreted as having a unique referent and overtly definite or indefinite Catalan singular nominals in the same discourse contexts. As illustrated in (1), the Russian sentence (1a) may be translated into Catalan (1b) (and English) in eight different ways, depending on the combination of articles, which shows the complexity and variability of interpretations that bare nominals may have in languages without articles.

- (1) Context: We entered the house.

Russian

a. Devočka čitala knigu v kresle.  
 girl.NOM read.IMP.PST book.ACC in armchair.PREP

Catalan

b. La / una noia llegia el / un llibre a la / una  
 the.Fa.F girl read.IMP.PST the.M a.M book in the.F a.F  
 butaca.  
 armchair  
 'The/a girl was reading the/a book in the/an armchair.'

The comparison of the two languages suggests that, unlike Catalan NPs preceded by a definite article, Russian bare nominals do not necessarily get a uniqueness interpretation. The main hypothesis is that bare noun phrases in Russian are interpreted as having a unique referent when it is part of the common ground of the interlocutors that a given situation is supposed to contain exactly one referent that satisfies the description expressed by the NP, while in Catalan this kind of interpretation is encoded in grammar by means of an overt

definite article, and the absence of uniqueness is signalled by an overt indefinite article.

Based on an experimental study of the interpretation of NPs in Catalan and Russian, we propose that it is irrelevant to talk about (in)definiteness as a binary grammatical category in the absence of articles in a language such as Russian, even though an NP may be interpreted by the speakers as having a unique or a non-unique reference. Thus, it can be concluded that a definite (unique) interpretation in languages with articles is related to the presence/absence of a definite article, while it has other sources in Russian (related to pragmatic factors). This outcome is in line with the classical proposal of Pardee (1987), who associates uniqueness/maximality<sup>1</sup> with the contribution of the definite article itself, and not of an iota operator, as claimed in Chierchia (1998), Dayal (2004) and Coppock and Beaver (2015). It also gives support to the recent proposals by Šimik and Demian (2020) and Seres and Borik (2021) for Russian, who claim that the absence of articles is translated into the absence of a definiteness-related semantics (i.e., a uniqueness interpretation).<sup>2</sup>

This paper is structured as follows. First, in Section 2, we briefly revise the semantic theory of definiteness as uniqueness with respect to languages with articles, discuss an alternative proposal and its outcomes for languages without articles. Then, in Section 3, we review possible sources of uniqueness interpretation regarding languages without articles. In Section 4, we present an experimental study that highlights the interpretative differences of NPs in Catalan and Russian.

---

<sup>1</sup> For plural nominals uniqueness is reformulated as maximality, which is conceived as a reference to a maximal individual in the domain (Sharvy 1980; Link 1983). This maximal individual is picked out by the definite article in languages with articles.

<sup>2</sup> This claim goes against Dayal (2004, 2011, 2017) who posits that Russian bare nominals are interpreted either as definites or as generics.

## 2. Theoretical background

There has been considerable research, both in linguistics and in philosophy, regarding definite and indefinite descriptions in natural language (Frege 1892; Russell 1905; Christopherson 1939; Strawson 1950; Hawkins 1978, among others).<sup>3</sup> It is important to emphasize that this research has been mainly focused on languages with overt articles; however, the theory of descriptions could have been very different if it had been elaborated based on a language without articles (Ludlow and Segal 2004; Dayal 2017, among others).

A standard view on definiteness in formal semantics is based on the so-called theory of uniqueness (Russell 1905; Strawson 1950; Chierchia 1998; Dayal 2004, among others). Uniqueness is understood as the existence in the extension of an NP of exactly one referent that satisfies the descriptive content of this NP in a given context.<sup>4</sup> A uniqueness interpretation means that the nominal is construed in the narrowest possible domain.

It is crucial to notice that in order to single out the referent of an NP the participants of communication need to rely on common knowledge (Hawkins 1978). This knowledge may arise from the previous mention of the referent (familiarity) (Heim 1982), but also from a more general shared knowledge of the participants of communication regarding the situation and the world (identifiability) (Lyons 1999).

The property of being unique is standardly considered to be a presupposition, associated with a definite description (Heim 1991; von Stechow 2004; Elbourne 2005, 2013).<sup>5</sup> Thus, if we compare the sentences in (2), it is clear that (2a) is about

---

<sup>3</sup> For an overview of different approaches to descriptions the reader is referred to Ludlow (2018).

<sup>4</sup> Therefore, uniqueness of a referent entails its existence. The discussion of the presupposition of existence, associated with definite descriptions, is out of the scope of this paper.

<sup>5</sup> Notice that for Russell (1905) the uniqueness component of a definite description is understood as an entailment.

a contextually unique book that both the interlocuters are aware of, while (2b) may have more than one possible referent.

- (2) Catalan
- a. *He llegit el llibre.*  
 AUX.1SG read the.M book  
 'I have read the book.'
- b. *He llegit un llibre.*  
 AUX.1SG read a.M book  
 'I have read a book.'

In formal semantics, the definite article denotes a function from predicates (the denotation of a common noun, type  $\langle e, t \rangle$ ) to individuals (type  $e$ ), which corresponds to the type-shifting iota operator (Partee 1987). The meaning of the definite article can be formally represented as follows (Heim 2011: 998, 4):

- (3)  $[[\text{the}]] = \lambda P: \exists x \forall y [P(y) \leftrightarrow x = y]. \iota x. P(x)$ ,  
 where  $\iota x$  abbreviates 'the unique  $x$  such that'

In this approach, the uniqueness of the referent of a definite description follows from the meaning of the article itself, and, thus, would not be expected to be present in languages without articles.<sup>6</sup> The uniqueness component of the meaning associated with the presence of a definite article in Catalan has been tested in the empirical study, presented in Section 4. Russian bare nominals, on the contrary, do not necessarily get interpreted as unique in the same contexts, as was illustrated in (1).

Articles, belonging to a wider category of determiners, are considered to express a domain restriction over their NP (von Stechow 1994; Gillon 2006, among many others). The unique-

---

<sup>6</sup> Chierchia (1998), Dayal (2004), i.a., propose a different approach, suggesting that type-shifting principles are universal, and, thus, the iota operator should be present in languages without articles, even though it is not lexicalised. However, there is no solid empirical evidence for postulating an iota operator in languages without articles.



ness reading, encoded by the definite article, represents the narrowest domain restriction: there is only one referent satisfying the description under the given circumstances, and that is the strongest statement to which the speaker can commit (following Grice's (1975) maxim of quantity). Otherwise, the speaker would have used a less strong expression with a wider domain, e.g., an indefinite description.

However, there is an alternative approach to descriptions, which rejects the uniqueness claim associated with definite descriptions, postulating that the only relevant distinction between NPs preceded by a definite or an indefinite article is pragmatic (Ludlow 2018).<sup>7</sup> From this perspective, overt articles can be even considered redundant as the discourse context should be sufficient to determine whether an NP is definite or not (Hawkins 2004).<sup>8</sup> The claim of the absence of a semantic difference between definite and indefinite descriptions may seem too radical for languages with articles but could be valid for languages which do not express a definite/indefinite distinction (at least in the overt morphosyntax).

The hypothesis that is sustained in this work is that in languages without articles there is no binary grammatical category of (in)definiteness. In fact, there is one logical element (a bare nominal) which may give rise to interpretations similar to the ones of definite and indefinite descriptions in languages with articles. The interpretation of this element depends on the discourse context and on the common ground of the interlocutors, i.e., the interpretation is achieved through pragmatic mechanisms.

In languages without articles, this element, expressed by a bare nominal with an *e* type denotation, may be derived by means of a choice function, as proposed in Seres and Borik (2021).

---

<sup>7</sup> See Szabó (2000) and Ludlow and Segal (2004) for details; this idea was also considered in Heim (1982).

<sup>8</sup> See Leiss (2007) who explains the morphological underspecification of nominals in languages without articles by cognitive economy.

(4)  $f_{CH} \{x: P(x)\}$

According to Reinhart (1997) and Winter (1997), choice functions that map any non-empty set onto an element of that set.<sup>9</sup> Thus, it is a function of type  $\langle \langle e, t \rangle, e \rangle$ , which applies to the property (of type  $\langle e, t \rangle$ ) and yields an individual (of type  $e$ ) that has this property; this type-shift is assumed to be covert in Russian.<sup>10</sup> The crucial advantage of the semantic derivation of argumental NPs in Russian by means of a choice function (not an iota operator as in languages with articles) is that it does not imply any uniqueness or familiarity of the referent, which are components of meaning usually associated with definiteness.

Indeed, Russian bare NPs may show properties that indefinite nominals have in languages with articles, for instance, they may take different scopes; may be used in opacity contexts; may be used in existential sentences; may introduce discourse referents; two identical non-coreferential NPs may be used in the same sentence.<sup>11</sup> At the same time, a definite (unique) interpretation of a bare nominal is not excluded either. In the following section, we review some of the factors that influence the rise of the uniqueness interpretation on a bare nominal in Russian.

### 3. Uniqueness in languages without articles

To start with, it is important to notice that the classical theory of definiteness as uniqueness was elaborated for languages with articles, as presented in the previous section, may still be

---

<sup>9</sup> As the set is non-empty, the existence claim holds for nominals derived by means of a choice function.

<sup>10</sup> Choice functions were first proposed to represent the semantics of indefinite NPs in languages with articles in Reinhart (1997), Winter (1997) and Krazter (1998). See also proposals by Yanovich 2005, Geist 2008, Šimik 2021 for indefinites in Slavic languages and Borik and Espinal (2020) for a choice-function analysis of definite kinds in Russian.

<sup>11</sup> See Serés (2020) for examples and details regarding these properties of bare NPs in Russian.

applicable to languages without articles if one supposes that uniqueness is still expressed, but just with different formal means (Abraham et al. 2007). Indeed, there is a general assumption in the linguistic literature on Slavic languages (Galkina Fedoruk 1963; Fursenko 1970; Pospelov 1970; Nessel 1999, among many others) that, although some languages do not have a lexical category to express (in)definiteness, this grammatical category is present in the language and there may be different means to express it. That is, that definiteness does not only depend on the discourse context.

Moreover, it is clear that a certain contrast between a definite and an indefinite interpretation is available for speakers of Russian, which is reflected in the way bare NPs are translated into languages with articles, as shown in (5a,b).

- (5) Russian
- a. Sobaka *zdes*’.  
dog.NOM here  
‘The dog is here.’
  - b. *Zdes*’ sobaka.  
here dog.NOM  
‘A dog is here.’

In Russian, there are several formal ways of conveying interpretations similar to definite or indefinite ones in languages with articles. They include lexical means (determiners, quantifiers, demonstratives), morphological (alternations of the verbal aspect and the case of nominals), prosodic (deaccentuation of ‘discourse old’ given information) and syntactic (linear word order alternations, as illustrated in (5a,b) (Seres et al. 2019). Nevertheless, as Borik et al. (2020) posit, none of these means is strong enough to be considered the trigger of a uniqueness reading, comparable to languages which have articles as a lexical class. There is no single grammatical means that could be equivalent to the definite article (corresponding to the iota operator) in Russian in all possible cases.

All in all, a uniqueness or non-uniqueness interpretation of bare nominals in Russian depends on pragmatic factors, related to the background knowledge of the participants and the discourse context, but not any linguistic means. A bare NP is interpreted as unique if it is in the common ground that there is exactly one referent satisfying this description in the situation of communication. There are factors that enhance the possibility for a bare nominal to be interpreted as unique.

First and foremost, it is the ‘ontological’ uniqueness of the referent, which holds for entities like *solnce* ‘the sun’, *zemlja* ‘the earth’, etc. (Seres and Borik 2021). It is particularly easy for the interlocutors to agree on the uniqueness of the referent if they share the relevant background knowledge, i.e., it is not the bare nominal that conveys uniqueness but the speakers’ knowledge about the referent. For instance, in (6), *luna* is interpreted as ‘the moon’, not ‘a moon’, since the moon is the Earth’s unique natural satellite, and the interlocutors are most probably aware of that. However, in a narrative about other planets which have more than one natural satellite the bare nominal in question may have a non-unique interpretation.

- (6) Luna            *svetit jarko.*  
 moon.NOM shines brightly  
 ‘The moon is shining brightly.’

Another factor that appears to contribute to an agreement on the uniqueness of the referent rather straightforwardly is D-linking (discourse-linking) of the nominal. This phenomenon was introduced by Pesetsky (1987) to describe constituents anchored to another one in the preceding discourse or in extralinguistic context. According to Dyakonova (2009: 73), a constituent is D-linked if (i) it has been explicitly mentioned in the previous discourse (direct anaphora), as illustrated in (7), (ii) it is situationally given by being physically present at the moment of communication (situational definiteness), as illustrated in (8a,b), or (iii) it can be easily inferred from the context

by being in the set relation with some other entity or event figuring in the preceding discourse (associative anaphora/bridging), as illustrated in (9).

- (7) *Nedavno u menya pojavilsja ščenok. Malyš ešče ne  
not.long.ago at me appeared puppy little.one yet not  
umeet lajat' no uže znaet svoë imja.  
can bark but already knows own name  
'Not long ago I got a puppy. The little one cannot bark yet but  
knows its own name.'*

- (8) a. Context: at a table  
*Peredaj mne sol', požalujsta!  
pass me salt please  
'Pass me the salt, please.'*

- b. Context: after a game.  
*Kto pobeditel'?  
who winner  
'Who is the winner?'*

- (9) *V novom ofise ja sela za komp'juter. Monitor byl  
in new office I sat.down at computer monitor was  
starymi tusklym.  
old and dim  
'In the new office I sat down at a computer. The monitor was  
old and dim.'*

In (7), the unique reference of *malyš* 'the little one' is established from the immediately preceding context through anaphoric anchoring to *ščenok* 'puppy', introduced in the previous sentence. In (8a), the immediate situation restricts the domain, thus, the listener understands that the referent of *sol'* 'salt' is the one present in the situation. In (8b), it is the general knowledge that a game would normally have one winner that establishes uniqueness of *pobeditel'* 'winner'; however, it cannot be excluded that there was no winner, or the game ended in a draw and there were two winners. In (9), *monitor* 'monitor'

would get a unique reference as ‘the monitor of the computer in the new office’ considering that a computer would typically have one monitor, but it cannot be excluded that there is more than one. Following Seres and Borik (2021), who postulate the absence of uniqueness for bare nominals in Russian, it can be suggested that, in (10), the appearance of a second referent in the following context would cancel the uniqueness implicature (that rises from the general knowledge of the participants of the communication) but would not cause unacceptability, as a bare nominal does not trigger uniqueness effects (the narrowest domain restriction) and is compatible with the whole range of domain restrictions.

- (10) *V novom ofise ja sela za komp'juter. Monitor byl starym i tusklym. Drugoi monitor, pri ètom, byl old and dim second monitor at this was supersovremennym, ja ne znala daze, kak ego vkladit'. super modern I not knew even how it turn.on*  
 ‘In the new office I sat down at a computer. The monitor was old and dim. The other monitor, at the same time, was super modern, I did not even know how to turn it on.’

The prediction that Seres and Borik (2021) make is that in languages with articles, the definite description, equivalent to the bare nominal *monitor* ‘monitor’, would trigger the construal of the narrowest possible domain, being the strongest statement that the speaker can commit to. The appearance of a second referent in that case would cause unacceptability, which can be accounted for as a violation of a presupposition of uniqueness (if one considers uniqueness to be a presupposition contributed by the definite article). See example (11) in Catalan, which is equivalent to (10) in Russian.<sup>12</sup>

---

<sup>12</sup> The (un)acceptability of sentences, such as the ones in (10) and (11) is yet to be tested experimentally.

- (11) *A la nova oficina em vaig asseure en un.*  
 in the new office refl.1sg aux.pst.1sg sit.down in a  
*ordinador. El monitor era vell i fosc. ?? Unaltre*  
 computer the monitor was old and dim an other  
*monitor, al mateix temps, era supermodern, ni tan*  
 monitor at.the same time was supermodern not  
*sols sabia com engegar-lo.*  
 even knew.impf.1sg ho turn.on it

In the next section, we present empirical evidence regarding the presence/absence of uniqueness interpretation in bare NPs in Russian and non-bare NPs in Catalan. The experimental study that we carried out focused on the interpretation of nominals in contexts similar to the one presented in (9), i.e., Catalan and Russian NPs were tested in the contexts that could potentially, but not obligatorily, contain a unique referent.

#### 4. Experimental study

The main goal of the experimental study was to compare the interpretation of NPs in the same contexts in Catalan and Russian, and to see whether these nominals convey uniqueness. Following from the previous theoretical discussion, we will show that definite NPs in Catalan are interpreted as having a unique referent, while indefinite NPs may have more than one possible referent, that is, they do not convey uniqueness. As for bare nominals in Russian, they are compatible with both, a uniqueness and a non-uniqueness interpretation, and some contexts may favour the interpretation of the referent as unique, based on the world view and the common ground of the participant of communication.

There has not been much of experimental work with respect to the interpretation of bare NPs in Slavic vs. languages belonging to other groups, with a notable exception of Šimik and Demian (2020) who compare interpretations of singular and plural nominals in topic position in Russian and German. The main outcome of their work is that the perceived definite-

ness (uniqueness) of Russian singular bare nominals may be overridden, while it is not possible for definite nominals in German. Our experimental study aims at confirming these findings and providing more empirical cross-linguistic data based on a contrastive study between Russian and Catalan.

#### **4.1. Experimental design**

Our experimental study consisted of three surveys. In Survey 1, the interpretation of Catalan definite NPs, in Survey 2, the interpretation of Catalan indefinite NPs, and in Survey 3, the interpretation of Russian bare NPs was tested in the same contexts.

Based on the literature, our prediction for Survey 1 was that with an overt definite article the nominal is interpreted as unique, regardless of the context. As for Survey 2, it was predicted that the presence of an indefinite article would be signalling non-uniqueness for the speakers. That is, the statements claiming the uniqueness of the referent, expressed by an indefinite description, were expected to be rated rather low.

The main prediction for Survey 3 was that a bare nominal may have either a uniqueness or a non-uniqueness interpretation and some contexts may favour one of the two interpretations, according to how speakers imagine a typical situation involving this referent to be. Thus, the results of the latter survey were expected to present more variability throughout the contexts and among the participants, as compared to the results of Survey 1 and Survey 2.

#### **4.2. Participants**

A total of 228 Catalan (96 for Survey 1 and 132 for Survey 2) and 100 Russian native speakers participated in the experimental study performed online using Alchemer software. Demographic information was collected from a sociolinguistic questionnaire administered right before the study that in-



quired about the participants' age, sex and level of studies, the place where participants were born and currently live, as well as how much they use their native language in their daily life. The sociolinguistic information, however, did not show any significant effect on the results.

### 4.3. Test items

The test items were initially taken from the Russian Web Corpus (ruTenTen) on SketchEngine and were slightly altered (shortened) to be more uniform. Each test item contains a brief preceding context, describing a situation and a following sentence with a preverbal non-anaphoric bare singular nominal, which is expected to be present in a given situation and whose uniqueness may be inferred from the situation. That is, the NPs are novel in the (narrow) discourse, but presumably are not novel in the common ground, being topics.

It is crucial to point out that the NPs in the experimental study are in leftmost/preverbal position, which is considered to be the topic position for Russian (Geist 2010; Jasinskaja 2014, among others). According to the classical view, bare NPs as topics obligatorily receive a definite interpretation in articleless Slavic languages (Geist 2010). However, this view is challenged in Seres et al. (2019) and Borik et al. (2020), who provide experimental evidence that topicality indeed strongly increases the probability for a bare nominal to receive an interpretation comparable to a definite one (for languages with articles), but it is not always a sufficient condition. Moreover, as for topical NPs, their perceived 'definiteness' may be due to the givenness/familiarity, not necessarily uniqueness, of a referent.<sup>13</sup>

The contexts that were used for the study describe situations that do not necessarily involve a unique referent but could also be perfectly compatible with there being only one

---

<sup>13</sup> Notice that topicality strongly favours a definite interpretation cross-linguistically (Reinhart 1981, Erteschik-Shir 2007, i.a.)

referent. Here is an example of an item taken from the corpus (we use a bare nominal in the English translation in order to reflect the Russian original):

- (12) *Éto byl samyj populjarnyj blog v gruppe, no čislo podpisčikov stalo rezko sokraščt'sja, kogda blog stal platnym. Avtor prodal ego za 10 tysjač evro.*

That was the most popular blog in the group, but the number of subscribers started decreasing sharply when the blog became paid. *Author* sold it for 10 thousand euros.'

The contexts and the NPs, whose interpretations were tested in the experiment, were as follows: *popular blog – author; local shopping centre – guard; school trip – teacher; butchery – butcher; office – manager; private company – programmer; ambulance – nurse.*

The items were translated into Catalan to create two surveys: with nominals preceded by a definite article (Survey 1) and with nominals preceded by an indefinite article (Survey 2).

#### 4.4. Procedure

Participants were asked to read a short description of a situation (context) and a sentence containing a singular definite NP (Survey 1 for Catalan), an indefinite NP (Survey 2 for Catalan) and a bare nominal (Survey 3 for Russian) in subject position. After that, the participants had to mark on the scale (from “no” to “yes”) whether they agreed or disagreed with the statement, regarding the uniqueness of the referent in the given context (as shown in Figures 1, 2 and 3 for the three surveys, respectively). The statement to be evaluated was highlighted in bold type. There were seven contexts (as mentioned above) which were presented twice, combined either with a critical statement or a filler statement.<sup>14</sup> That is, the participants had to evaluate 14 statements, which were randomized.

<sup>14</sup> Fillers were not taken into consideration for the analysis of the results.

#### 4.5. Example items

Figures 1, 2 and 3 show what the experimental items looked like in each survey. The English translation is provided after each figure in (13), (14) and (15), respectively.

*Anit van robar en un centre comercial del nostre veïnat. Els lladres no van tenir problemes per entrar i recollir tots els diners. El guàrdia de seguretat estava mirant la televisió i no va sentir res.*

Entenc que era l'únic guàrdia de seguretat que hi havia dins del centre comercial. \*

no  sí

**Figure 1**

Example item of Survey 1 (Catalan)

- (13) 'Last night, a commercial center in the neighbourhood was burgled. The thieves didn't have any problem to enter and take all the money. *The security guard* was watching TV and didn't hear anything. **I understand that it was the only security guard who was inside the commercial center.**'

*Gairebé immediatament, l'ambulància va arribar al lloc dels esdeveniments. Una infermera va demanar que li donessin alguna cosa per aturar l'hemorràgia: una corbata o una bufanda.*

Entenc que era l'única infermera que hi havia al lloc dels esdeveniments. \*

no  sí

**Figure 2**

Example item of Survey 2 (Catalan)

- (14) 'Almost immediately, the ambulance arrived at the place of the events. *A nurse* asked for something to stop the bleeding: a tie or a scarf. **I understand that she was the only nurse at the place of the events.**'

*На моих глазах убили молодого и красивого быка. Мясник занёс кинжал над его шеей и ударил. Бык как подкошенный упал на брюхо.*

**В убийстве животного принимал участие один мясник.\***



**Figure 3**

Example item of Survey 3 (Russian)

- (15) 'In front of my eyes, a young and beautiful bull was killed. Butcher pointed the knife at his neck and hit. The bull collapsed belly down. **Only one butcher took part in the slaughter of the animal.**'

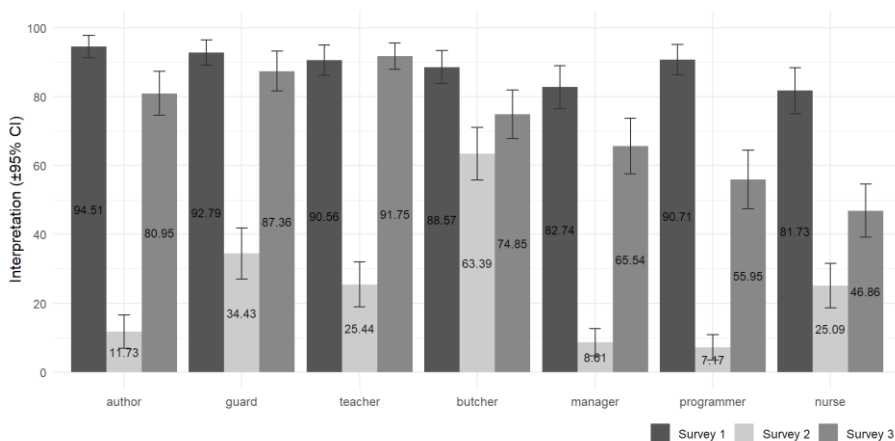
#### 4.6. Results

Figure 4 presents the results of Surveys 1, 2 and 3. In Survey 1, the speakers of Catalan give very high acceptability to the interpretation of the nominal preceded by a definite article as unique ( $M = 88.80$ ,  $SD = 24.55$ ); in all cases, the acceptability is higher than 81 %, thus, a strong tendency to interpret a definite NP as unique can be seen.

As can be seen for Survey 2, the rating of statements claiming the uniqueness of the referent is, indeed, very low, as compared to Survey 1 ( $M = 25.12$ ,  $SD = 39.03$ ). The only outlier is the context with the butcher where the acceptability of a uniqueness reading is relatively high (63.39 %), this might be due to an experimental error or the influence of a context (the way participants imagined a typical butchery: whether there is normally only one butcher working or not) or to some other factor linguistic or extra-linguistic factor that we failed to detect. Nonetheless, the preference for the uniqueness interpretation is significantly lower than the one with an overt definite article in Survey 1 (88.57 %).

Finally, as shown for Survey 3, the preference for a uniqueness reading for bare nominals in Russian varies signifi-

cantly: from 91.75 % to 46.86 % ( $M = 71.89$ ,  $SD = 38.10$ ); a clear preference for a uniqueness interpretation is found in the first four contexts (*popular blog – author; local shopping centre – guard; school trip – teacher, butchery – butcher*): more than 70 %, while the other three contexts (*office – manager; private company – programmer; ambulance – nurse*) do not show any clear preference: 46–65 %.



**Figure 4**

Uniqueness interpretations in surveys 1, 2 and 3

A beta mixed-effects model was run with the value interpretation as the dependent variable. To fulfill the requirements of a model based on a beta regression, the response values were first divided by 100 (to obtain a 0-1 distribution), and then the two ends were replaced by very close values (0.0000001 for 0, and 0.9999999 for 1). The Survey was set as the fixed factor, and a random intercept was defined for both Subject and Item.

A significant effect was found for Survey,  $\chi^2(2) = 504.194$ ,  $p < .001$ , indicating that Catalan definite structures were significantly perceived as conveying more uniqueness than both Catalan indefinite structures (*Cohen's d* = 6.270,  $p < .001$ ) and Russian bare structures ( $d = 1.821$ ,  $p < .001$ ). Among the latter, the Russian structures were also seen as conveying more

uniqueness than the Catalan indefinite structures ( $d = 4.449$ ,  $p < .001$ ).

#### **4.7. Discussion**

Despite the limited number of contexts (only seven) used for this experimental study, the results of the comparison of the interpretations attributed to definite and indefinite nominals in Catalan vs. bare nominals in Russian are compatible with the theoretical claims exposed in Sections 2 and 3 of this paper. That is, the presence of an overt definite article can be related to the uniqueness construal of a nominals. As predicted, the referent is predominantly (higher than 81 %) interpreted as unique if there is an overt definite article in Survey 1, while with an overt indefinite article in Survey 2 prefer a non-unique interpretation: between 7.17 % and 34.43 %.

The results of Survey 1 and Survey 2 confirm the main hypothesis that speakers of Catalan as a language with articles rely on the overt article when it comes to interpreting a referent of an NP as unique or non-unique. The results of the interpretation of Russian bare nominals are not as straightforward and uniform (which is also an expected result).

As predicted, the uniqueness interpretation attributed to bare nominals in Survey 3 varies significantly, not being lower than 46.86 % for any context (with a mean of 71.89 %), which may be related to the topic position of the bare nominals under study. That is, the bare nominal is construed as given, and, possibly, contextually unique, as otherwise, the speaker would have mentioned other referents. It is clear though that the interpretation of bare nominal as unique or non-unique indeed depends on the context, i.e., a bare nominal itself does not encode uniqueness in Russian (and possibly, in other languages without articles).

As can be seen from Figure 4, certain contexts favour the uniqueness interpretation, while in other the referent of the bare nominal may be construed as either unique or non-

unique. This difference may be explained if discourse participants' beliefs about this situation are taken into consideration. Uniqueness interpretation arises when it is part of the common ground (in terms of Stalnaker 2002) of the participants of the linguistic interaction that there is only one unique referent in each situation, that is, if they imagine that a blog typically has one author,<sup>15</sup> there is one guard in a local shopping centre, there one teacher that accompanies a group of secondary school students, etc. As for the contexts where the referents got a lower uniqueness rating, there is no agreement on whether there is one or more than one referent, e.g., there may be more than one nurse in an ambulance or more than one programmer in a private company. In other words, the interpretation reflects how discourse participants imagine a prototypical situation.

## 5. Conclusions

All in all, we conclude that uniqueness in languages with articles (such as Catalan) and languages without articles (such as Russian) is encoded differently. The uniqueness of a referent is related to grammatical definiteness, that is, the presence of a definite article in Catalan. On the contrary, in Russian it is not relevant to postulate (in)definiteness as a binary grammatical category, related to a uniqueness or a non-uniqueness interpretation of a bare nominal. In languages without articles, there is a single logical element which may give rise to interpretations similar to the ones of definite and indefinite descriptions in languages with articles (conceived in this article in terms of uniqueness or its absence). Uniqueness (which can be viewed as a presupposition) is encoded semantically (by a definite article) in languages with articles (Catalan), while in languages without articles (Russian) its appearance is conditioned

---

<sup>15</sup> Notice that *author* is a relational noun. It is interpreted as unique in accordance with Grice's (1975) maxim of quantity. The hearer infers that it is the stronger statement the speaker can commit to.

pragmatically, thus, speakers' interpretation of one and the same bare nominal may vary. The pragmatic presupposition of uniqueness arises on bare NPs when it is part of the common ground that the situation contains no more than one entity being referred to.

## References

- Abraham, Werner, Elisabeth Stark, Elisabeth Leiss (2007). "Introduction". In: Elisabeth Stark, Elisabeth Leiss, Werner Abraham (eds.). *Nominal Determination: Typology, Context Constraints, and Historical Emergence*. Amsterdam – Philadelphia: John Benjamins Publishing Company, 1–20. DOI:10.1075/slcs.89.02 abr.
- Borik, Olga, Joan Borràs-Comes, Daria Seres (2020). "Preverbal (in)definites in Russian: An experimental study". In: Kata Balogh, Anja Latrouite, Robert D. Van Valin, Jr. (eds.). *Nominal Anchoring: Specificity, Definiteness, and Article Systems*. Berlin: Language Science Press, 51–80. Also available at <http://doi.org/10.5281/zenodo.4049681>. Accessed 04.07.2021.
- Borik, Olga, M. Teresa Espinal (2020). "Numberless kinds: Evidence from Russian". *Catalan Journal of Linguistics* 19: 231–260. Also available at <https://revistes.uab.cat/catJL/article/view/v19-borik-espinal/299-pdf-en>. Accessed 04.07.2021.
- Brun, Dina (2001). "Information structure and the status of NP in Russian". *Theoretical Linguistics* 27/2–3: 109–135. DOI: 10.1515/thli.2001.27.2-3.109.
- Chierchia, Gennaro (1998). "Reference to kinds across languages". *Natural Language Semantics* 6/4: 339–405. Also available at <https://link.springer.com/content/pdf/10.1023/A:1008324218506.pdf>. Accessed 04.07.2021.
- Christophersen, Paul (1939). *The Articles: A Study of Their Theory and Use in English*. Copenhagen: Einar Munksgaard.
- Coppock, Elizabeth, David Beaver (2015). "Definiteness and determinacy". *Linguistics and Philosophy* 38/5: 377–435. DOI: 10.1007/s10988-015-9178-8. Also available at <https://link.springer.com/content/pdf/10.1007/s10988-015-9178-8.pdf>. Accessed 04.07.2021.



- Cummins, George (1998). "Indefiniteness in Czech". *Journal of Slavic Linguistics* 6/2: 171–203. Also available at <<http://www.jstor.org/stable/24599695>>. Accessed 03.07.2021.
- Dayal, Veneeta (2004). "Number marking and (in)definiteness in kind terms". *Linguistics and Philosophy* 27/4: 393–450. DOI: 10.1023/b:ling.0000024420.80324.67. Also available at <<https://link.springer.com/content/pdf/10.1023/B:LING.0000024420.80324.67.pdf>>. Accessed 04.07.2021.
- Dayal, Veneeta (2011). "44. Bare noun phrases". In: Klaus von Heusinger, Claudia Maienborn, Paul Portner (eds.). *Semantics: An International Handbook of Natural Language Meaning*. Vol. 2. Berlin – Boston: De Gruyter Mouton, 1088–1109. DOI: 10.1515/9783110255072.1088.
- Dayal, Veneeta (2017). "Determining (in)definiteness in the absence of articles". In: Vera Hohaus, Wanda Rothe (eds.). *Proceedings of Triple A 3: Fieldwork Perspectives on the Semantics of African, Asian and Austronesian Languages*. DOI: 10.15496/publikation-15341. Available at <<https://publikationen.uni-tuebingen.de/xmlui/handle/10900/73437>>. Accessed 04.07.2021.
- Dryer, Matthew S. (2013a). "Definite articles". In: Matthew S. Dryer, Martin Haspelmath (eds.). *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Available at <<https://wals.info/chapter/37>>. Accessed 03.07.2021.
- Dryer, Matthew S. (2013b). "Indefinite articles". In: Matthew S. Dryer and Martin Haspelmath (eds.). *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Available at <<https://wals.info/chapter/38>>. Accessed 03.07.2021.
- Dyakonova, Marina (2009). "A phase-based approach to Russian free word order". LOT Dissertation series 230. Utrecht: LOT. Available at <[https://www.lotpublications.nl/Documents/230\\_fulltext.pdf](https://www.lotpublications.nl/Documents/230_fulltext.pdf)>. Accessed 04.07.2021.
- Elbourne, Paul (2005). *Situations and Individuals*. Cambridge, MA: MIT Press.
- Erteschik-Shir, Nomi (2007). *Information Structure: The Syntax-discourse Interface*. Oxford: Oxford University Press.
- Erteschik-Shir, Nomi (2013). *Definite Descriptions*. Oxford: Oxford University Press. DOI:10.1093/acprof:oso/9780199660193.001.0001.

- von Fintel, Kai (1994). *Restrictions on Quantifier Domains*. PhD dissertation, University of Massachusetts Amherst.
- von Fintel, Kai (2004). "Would you believe it? The king of France is back! (Presupposition and truth-value intuitions)". In: Anne Bezuidenhout, Marga Reimer (eds.). *Descriptions and Beyond: An Interdisciplinary Collection of Essays on Definite and In-definite Descriptions*. Oxford: Oxford University Press, 315–341.
- Frege, Gottlob (1892). "Über Sinn und Bedeutung". *Zeitschrift für Philosophie und philosophische Kritik* 100: 25–50. Available at <[http://www.deutschestextarchiv.de/book/view/frege\\_sinn\\_1892?p=11](http://www.deutschestextarchiv.de/book/view/frege_sinn_1892?p=11)>. Accessed 04.07.2021.
- Fursenko, Diana I. (1970). "Porjadok slov kak odno iz sredstv vyraženiya neopredelënnosti /opredelënnosti imën suščestvitel'nyx". *Russkij Jazyk za Rubežom* 4: 68–72.
- Galkina-Fedoruk, Evdokia M. (1963). *Vyraženie neopredelënnosti v russkom jazyke neopredelënnymi mestoimenijami i narečijami*. Moskva: Moskovskij gosudarstvennyj universitet.
- Geist, Ljudmila (2008). "Specificity as referential anchoring: Evidence from Russian". In: Atle Grønn (ed.). *Proceedings of Sinn und Bedeutung 12*. Oslo: ILOS, 151–164.
- Geist, Ljudmila (2010). "Bare singular NPs in argument positions: Restrictions on indefiniteness". *International Review of Pragmatics* 2/2: 191–227. DOI: 10.1163/187731010x528340.
- Gillon, Carrie (2006). *The Semantics of Determiners: Domain Restriction in Skwxwú7mesh*. PhD dissertation, The University of British Columbia.
- Grice, H. Paul (1975). "Logic and conversation". In: Peter Cole, Jerry L. Morgan (eds.). *Syntax and Semantics*. Vol. 3. *Speech Acts*. New York: Academic Press, 41–58.
- Hawkins, John A. (1978). *Definiteness and Indefiniteness: A Study in Reference and Grammaticality Prediction*. London: Routledge.
- Hawkins, John A. (2004). *Efficiency and complexity in Grammars*. Oxford: Oxford University Press. DOI: 10.1093/acprof:oso/9780199252695.001.0001. Also available at <<https://dlib.scu.ac.ir/bitstream/Hannan/238543/1/0199252688.pdf>>. Accessed 04.07.2021.
- Heim, Irene (1982). *The Semantics of Definite and Indefinite Noun Phrases*. PhD dissertation, Amherst, MA: UMass. Available at <[https://scholarworks.umass.edu/dissertations/AAI\\_8229562](https://scholarworks.umass.edu/dissertations/AAI_8229562)>. Accessed 04.07.2021.

- Heim, Irene (1991). "Artikel und Definitheit". In: Armin von Stechow, Dieter Wunderlich (eds.). *Semantik: Ein internationales Handbuch der Zeitgenössischen Forschung*. Berlin: Walter de Gruyter, 487–535.
- Heim, Irene (2011). "41. Definiteness and indefiniteness". In: Klaus von Heusinger, Claudia Maienborn, Paul Portner (eds.). *Semantics: An International Handbook of Natural Language Meaning*. Vol. 2. Berlin – Boston: De Gruyter Mouton, 996–1025. DOI:10.1515/9783110255072.996.
- Jasinskaja, Katja (2016). "Information structure in Slavic". In: Caroline Féry, Shinichiro Ishihara (eds.). *The Oxford Handbook of Information Structure*. Oxford: Oxford University Press. DOI:10.1093/oxfordhb/9780199642670.013.25.
- Kratzer, Angelika (1998). "Scope or pseudoscope? Are there wide-scope indefinites?" In: Susan Rothstein (ed.). *Events and Grammar*. Dordrecht: Springer, 163–196. DOI: 10.1007/978-94-011-3969-4\_8.
- Leiss, Elisabeth (2007). "Covert patterns of definiteness/indefiniteness and aspectuality in Old Icelandic, Gothic, and Old High German". In: Elisabeth Stark, Elisabeth Leiss, Werner Abraham (eds.). *Nominal Determination*. Amsterdam – Philadelphia: John Benjamins, 73–102. DOI:10.1075/slcs.89.06lei.
- Link, Godehard (1983). "The logical analysis of plurals and mass terms: A lattice-theoretical approach". In: Rainer Bäuerle, Christoph Schwarze, Arnim von Stechow (eds.). *Meaning, Use and Interpretation of Language*. Berlin: De Gruyter, 302–323.
- Longobardi, Giuseppe (2001). "How comparative is semantics? A unified parametric theory of bare nouns and proper names". *Natural Language Semantics* 9/4: 335–369. DOI: 10.1023/a:1014861111123. Also available at <<https://link.springer.com/content/pdf/10.1023/A:1014861111123.pdf>>. Accessed 03.07.2021.
- Ludlow, Peter (2018). "Descriptions". In: Edward N. Zalta (ed.). *The Stanford Encyclopedia of Philosophy*. Available at <<https://plato.stanford.edu/entries/descriptions/>>. Accessed 03.07.2021.
- Ludlow, Peter, Gabriel Segal (2004). "On a unitary semantical analysis for definite and indefinite descriptions". In: Marga Reimer, Anne Bezuidenhout (eds.). *Descriptions and Beyond*. Oxford: Oxford University Press, 420–437.
- Lyons, Christopher (1999). *Definiteness*. Cambridge: Cambridge University Press. DOI:10.1017/cbo9780511605789.

- Nesset, Tore (1999). "The realization of (in)definiteness in Russian". *Poljarnyj Vestnik* 2: 84–109.
- Partee, Barbara H. (1987). "Noun phrase interpretation and type-shifting principles". In: Jeroen Groenendijk, Martin Stokhof (eds.). *Studies in Discourse Representation Theory and the Theory of Generalized Quantifiers*. Dordrecht: Foris, 115–143.
- Pesetsky, David (1987). "Wh-in-situ: Movement and unselective binding". In: Eric Reuland, Alice G. B. ter Meulen (eds.). *The Representation of (In)definiteness*. Cambridge, MA: MIT Press, 98–129.
- Pospelov, Nikolaj (1970). "O sintaksičeskom vyražennii kategorii opredelënnosti – neopredelënnosti v sovremennom russkom jazyke". In: Timofej Lomtev, Aleksandra Kamynina (eds.). *Issledovanija po sovremennomu russkomu jazyku*. Moskva: Izdatel'stvo Moskovskogo universiteta, 182–189.
- Reinhart, Tanya (1981). "Pragmatics and linguistics: An analysis of sentence topics". *Philosophica* 27/1: 53–94. Also available at <<https://www.philosophica.ugent.be/wpcontent/uploads/fulltext/s/27-4.pdf>>. Accessed 05.07.2021.
- Reinhart, Tanya (1997). "Quantifier scope: How labor is divided between QR and choice functions". *Linguistics and Philosophy* 20: 335–397. DOI: 10.1023/A:1005349801431.
- Russell, Bertrand (1905). "On denoting". *Mind* 14/4: 479–493. DOI: 10.1093/mind/XIV.4.479.
- Seres, Daria (2020). *The Expression of Genericity in Languages with and without Articles*. PhD dissertation, Universitat Autònoma de Barcelona. Also available at <<https://www.tdx.cat/handle/10803/670853>>. Accessed 04.07.2021.
- Seres, Daria, Joan Borràs-Comes, Olga Borik (2019). "Interplay between position and interpretation: An experimental study of Russian bare plurals". *Revue roumaine de linguistique* 64/2: 163–177. Also available at <<https://www.lingv.ro/images/RRL%202%202019%20%2005-Seres.pdf>>. Accessed 04.07.2021.
- Seres, Daria and Olga Borik (2020). "Definiteness in the absence of uniqueness: The case of Russian". In: Andreas Blümel, Jovana Gajić, Ljudmila Geist, Uwe Junghanns, Hagen Pitsch (eds.). *Advances in Formal Slavic Linguistics 2018*. Berlin: Language Science Press, 339–363.
- Sharvy, Richard (1980). "A more general theory of definite descriptions". *The Philosophical Review* 89/4: 607–624. DOI: 10.2307/

2184738. Also available at <<https://www.jstor.org/stable/2184738>>. Accessed 04.07.2021.
- Šimík, Radek (2021). “Inherent vs. accidental uniqueness in bare and demonstrative nominals”. In: Andreas Blümel, Jovana Gajić, Ljudmila Geist, Uwe Junghanns, Hagen Pitsch (eds.). *Advances in Formal Slavic Linguistics 2018*. Berlin: Language Science Press, 365-391.
- Šimík, Radek, Christoph Demian (2020). “Definiteness, uniqueness, and maximality in languages with and without articles”. *Journal of Semantics* 37/3: 311–366. DOI: 10.1093/jos/ffaa002. Also available at <<https://academic.oup.com/jos/article/37/3/311/5848470>>. Accessed 04.07.2021.
- Stalnaker, Robert (2002). “Common ground”. *Linguistics and Philosophy* 25/5–6: 701–721. DOI:10.1023/a:1020867916902. Also available at <<https://link.springer.com/content/pdf/10.1023/A:1020867916902.pdf>>. Accessed 05.07.2021.
- Strawson, Peter Frederick (1950). “On referring”. *Mind* 59 (235): 320–344. Also available at <<http://www.jstor.org/stable/2251176>>. Accessed 04.07.2021.
- Szabó, Zoltán G. (2000). “Descriptions and uniqueness”. *Philosophical Studies* 101: 29–57. Also available at <<https://link.springer.com/content/pdf/10.1023/A:1026437211756.pdf>>. Accessed 04.07.2021.
- Winter, Yoad (1997). “Choice functions and the scopal semantics of indefinites”. *Linguistics and Philosophy* 20: 399–467. DOI: 10.1023/A:1005354323136.
- Yanovich, Igor (2005). “Choice-functional series of indefinites and Hamblin semantics.” In: Effi Georgala, Jonathan Howell (eds.). *SALT 15: Proceedings of the 15<sup>th</sup> Semantics and Linguistic Theory Conference*. Ithaca, NY: CLC Publications, 309–326. DOI:10.3765/salt.v15i0.2921.

## Appendix

### Full list of Catalan and Russian experimental items with English translation

Each experimental item contains an introductory context, a critical statement (i), and a filler statement (ii).

Catalan (Survey 1 and 2)

1. Gairebé immediatament, l'ambulància va arribar al lloc dels esdeveniments. La/una infermera va demanar que li donessin alguna cosa per aturar l'hemorràgia: una corbata o una bufanda.

(i) Entenc que era l'única infermera que hi havia al lloc dels esdeveniments.

(ii) Entenc que la infermera va arribar amb l'ambulància.

'Almost immediately, the ambulance arrived at the place of the events. The/a nurse asked for something to stop the bleeding: a tie or a scarf. (i) I understand that she was the only nurse at the place of the events. (ii) I understand that the nurse came in the ambulance.'

2. En aquesta empresa privada, els deutes salarials van començar a augmentar i els treballadors van començar a anar-se'n. La/una informàtica se'n va anar fa un any i mig. El director general li devia 47.000 euros.

(i) Entenc que era l'única informàtica que treballava a l'empresa.

(ii) Entenc que la informàtica se'n va anar perquè el director general li devia diners.

'In this private company the salary debts started to grow, and the workers started to leave. The/a programmer left a year and a half ago. The CEO owed her 47,000 euros. (i) I understand that she was the only programmer working at the company. (ii) I understand that the programmer left because the CEO owed her money.'

3. Anit van robar en un centre comercial del nostre veïnat. Els lladres no van tenir problemes per entrar i recollir tots els diners.

El/un guàrdia de seguretat estava mirant la televisió i no va sentir res.

- (i) Entenc que era l'únic guàrdia de seguretat que hi havia dins del centre comercial.
- (ii) Entenc que el guàrdia de seguretat no estava treballant.

'Last night, a commercial center in the neighbourhood was burgled. The thieves didn't have any problem to enter and take all the money. The/a security guard was watching TV and didn't hear anything. (i) I understand that he was the only security guard who was in the commercial center. (ii) I understand that the security guard was not working.'

4. Era el blog més popular de la banda, però el nombre de subscriptors va començar a disminuir dramàticament quan va deixar de ser gratuït. L'/un autor d'aquest blog l'ha venut per deu mil euros.

- (i) Entenc que era l'únic autor d'aquest blog.
- (ii) Entenc que era un dels blogs més populars.

'That was the most popular blog in the group, but the number of subscribers started decreasing sharply when the blog became paid. The/an author sold it for 10 thousand euros. (i) I understand that it was the only author of the blog. (ii) I understand that it was one of the most popular blogs.'

5. Solia viatjar amb tren de rodalies. Una vegada vaig veure una excursió escolar. Eren estudiants de secundària. Alguns jugaven a cartes, uns altres fumaven i deien paraulotes mentre la/una professora llegia un diari.

- (i) Entenc que era l'única professora que acompanyava els estudiants.
- (ii) Entenc que aquesta persona sempre viatjava amb el mateix tren de rodalies.

I used to travel by commuter train. Once I saw a school trip. They were secondary school students. Some of them were playing cards, others were smoking and swearing while the/a teacher was reading a newspaper. (i) I understand that she was the only teacher who ac-

companied the students. (ii) I understand that the person was traveling by commuter train.’

6. Van matar un toro jove i maco davant dels meus ulls. El/un carnisser li va posar un ganivet al coll i l’hi va clavar. El toro va caure a terra.

(i) Entenc que era l’únic carnisser que va matar el toro.

(ii) Entenc que era l’únic toro que van matar aquell dia.

‘In front of my eyes a young and beautiful bull was killed. The/a butcher pointed the knife at his neck and hit. The bull collapsed belly down. (i) I understand that he was the only butcher who killed the bull. (ii) I understand that it was the only bull killed on that day.’

7. La gent de l’oficina no té complexos, en absolut, cap complex. La/una responsable treu les pinces de la seva butxaca durant la reunió i comença a arrencar-se els pèls de la barbeta.

(i) Entenc que era l’única responsable que treballava a l’oficina.

(ii) Entenc que a l’oficina hi treballava molta gent.

‘The people at the office don’t have complexes, not one. The/a manager takes tweezers out of her pocket during the meeting and starts pulling out hairs from her chin. (i) I understand that she was the only manager working at the office. (ii) I understand that a lot of people worked at the office.’

### Russian (Survey 3)

1. Почти сразу к месту происшествия приехала бригада скорой помощи. Медсестра обратилась к окружающим с просьбой дать ей что-нибудь, чтобы остановить кровь.

(i) Это была единственная медсестра в бригаде скорой помощи.

(ii) Медсестра приехала вместе с бригадой скорой помощи.

‘Almost immediately, the ambulance arrived at the place of the events. *Nurse* asked people around her for something to stop the bleeding. (i) That was the only nurse in the ambulance crew. (ii) The nurse arrived with the ambulance crew.’



2. В этой довольно крупной частной компании начали расти долги по зарплате, стали увольняться сотрудники. Программист ушёл полгода назад, долг директора составил 47000 рублей.

- (i) Это был единственный программист в компании.
- (ii) Программист ушёл, потому что директор был должен ему денег.

'In this rather big private company the salary debts started to grow, and the workers started to leave. *Programmer* left a year and a half ago. The CEO owed him 47,000 roubles. (i) That was the only programmer working at the company. (ii) The programmer left because the CEO owed him money.'

3. Вчера вечером в нашем районе ограбили магазин. Грабители без особого труда взломали дверь и собрали все наличные. Охранник смотрел телевизор и ничего не слышал.

- (i) В магазине не было других охранников.
- (ii) Охранник не работал в этот день.

'Last night, a shop in the neighbourhood was burgled. The thieves didn't have any problem to enter and take all the cash. *Security guard* was watching TV and didn't hear anything. (i) There were no other security guards in the shop. (ii) The security guard was not working on that day.'

4. Это был самый популярный блог в группе, но число подписчиков стало резко сокращаться, когда блог стал платным. Автор продаёт его за 10 тысяч евро.

- (i) У блога был один единственный автор.
- (ii) Это был очень популярный блог.

'That was the most popular blog in the group, but the number of subscribers started decreasing sharply when the blog became paid. *Author* sold it for 10 thousand euros. (i) The blog had only one author. (ii) It was a very popular blog.'

5. Я часто ездила в электричке. Один раз наблюдала, как учительница едет на экскурсию со старшеклассниками – одни играют в карты, другие курят и матерятся в тамбуре, а она просто читает газету.

- (i) С группой школьников ехала одна учительница.  
(ii) Эти старшеклассники часто ездили в электричке.

I used to travel by commuter train. Once I saw how *teacher* was on a school trip with secondary school students - some of them were playing cards, others were smoking and swearing while she was just reading a newspaper. (i) She was the only teacher who accompanied the students. (ii) The school students regularly travelled by commuter train.'

6. На моих глазах убили молодого и красивого быка. Мясник занёс кинжал над его шеей и ударил. Бык как подкошенный упал на брюхо.

- (i) В убийстве животного принимал участие один мясник.  
(ii) В тот день убили только одного быка.

In front of my eyes a young and beautiful bull was killed. *Butcher* pointed the knife at his neck and hit. The bull collapsed belly down.

- (i) Only butcher took part in the killing of the bull. (ii) Only one bull was killed on that day.'

7. В офисе народ совсем, абсолютно без комплексов. Начальница достаёт из кармана щипчики во время собрания и начинает выщипывать волосы на подбородке.

- (i) В этом офисе одна единственная начальница.  
(ii) В офисе работало много людей.

The people at the office don't have complexes, not one. *Manager* takes tweezers out of her pocket during the meeting and starts pulling out hairs from her chin. (i) There is only one manager at the office. (ii) A lot of people worked at the office.

## Acknowledgements

We would like to thank Olga Borik for her comments and suggestions, related to the theoretical and the empirical part of the study. We are also grateful to the audience of ROLLING (Research on Language and Linguistics of the University of Rovira i Virgili) seminar held in May 2021, where an early version of this paper was presented. This study is supported by the Spanish MINECO (grant POID2020-112801GB-100), by the Generalitat de Catalunya (grant 2017SGR634) and an ICREA Academia fellowship awarded to the third author.

Daria Serés  
ORCID iD: 0000-0002-9044-8516  
Universitat Autònoma de Barcelona  
Departament de Filologia Anglesa/  
Centre de Lingüística Teòrica  
Carrer de la Fortuna  
08193 Bellaterra  
Spain  
daria.seres@uab.cat

Joan Borràs-Comes  
ORCID iD: 0000-0003-4619-9046  
Universitat Autònoma de Barcelona  
Departament de Filologia Catalana  
Carrer de la Fortuna  
08193 Bellaterra  
Spain  
joanborrascomes@gmail.com

M. Teresa Espinal  
ORCID iD: 0000-0002-8079-7253  
Universitat Autònoma de Barcelona  
Departament de Filologia Catalana/  
Centre de Lingüística Teòrica  
Carrer de la Fortuna  
08193 Bellaterra  
Spain  
teresa.espinal@uab.cat



**Definiteness in second language acquisition:  
Preliminary results regarding Indo-European  
and Afroasiatic languages**

CARLA FERRERÓS PAGÈS  
FRANCESC ROCA URGELL  
MARTA VILOSA SÁNCHEZ

*Received 01.11.2021,  
Accepted 01.12.2021.*

**Abstract**

This article describes and compares the way in which definiteness is expressed in Romance (Catalan and Spanish) and some Slavic and Afro-Asian languages. We present some difficulties concerning definite nominal expressions that speakers of Ukrainian, Egyptian Arabic and Amazigh as L1 face when learning Catalan or Spanish as a second language and we show that the acquisition of definite determiners is, in general, problematic regardless of the typological nature of the L1. We also indicate that these difficulties can be related to the emergence of different determiner layers in the higher functional field in the nominal domain during the acquisition process.

**Keywords**

definiteness, definite determiner, L2 acquisition, Romance languages, Slavic languages, Afro-Asian languages

## **Określoność w przyswajaniu języka drugiego: Wstępne przemyślenia dotyczące języków indoeuropejskich i afroazjatyckich**

### **Abstrakt**

Ten artykuł opisuje i porównuje sposób, w jaki wyrażona jest określoność w językach romańskich (katalońskim i hiszpańskim) oraz w niektórych językach słowiańskich i afroazjatyckich. Przedstawiamy pewne trudności dotyczące określonych wyrażeń nominalnych, z jakimi borykają się osoby mówiące po ukraińsku, egipskim arabskim i amazigh jako L1, ucząc się katalońskiego lub hiszpańskiego jako drugiego języka. Pokazujemy, że przyswajanie rodzajników określonych jest generalnie problematyczne, niezależnie od typologicznego charakteru L1. Wskazujemy również, że trudności te mogą być związane z pojawieniem się różnych warstw określników w wyższym polu funkcjonalnym w domenie nominalnej podczas procesu przyswajania.

### **Słowa kluczowe**

określoność, rodzajniki, przyswajanie L2, języki romańskie, języki słowiańskie, języki afroazjatyckie

### **1. Introduction**

This article addresses the issues of how definiteness is encoded in nominal expressions and how definite determiners develop in the process of acquiring a Romance language as a second language (L2). We are presenting the preliminaries of an ongoing research on the acquisition of definite determiners in Spanish and Catalan by speakers whose first languages (L1) belong to different linguistic families (Indo-European and Afro-Asiatic) and present notable differences among them concerning the realization of definiteness in nominal expressions. In this work we deal with Russian and Ukrainian (Indo-European and Slavic), Arabic (Afro-Asiatic and Semitic), Amazigh (Afro-

Asiatic), and Spanish and Catalan (Indo-European and Romance). The way in which these languages express definiteness in nominal constructions is different in each group (Slavic, Semitic, Amazigh and Romance), either by means of a very specific grammatical element such as the definite article, or by combining the noun with elements such as demonstratives or possessives, or without any grammatical element at all, in which case the definite interpretation is inferred from the discursive context.

The paper is organized as follows: first, we present the basis of the DP hypothesis, which has become fundamental in the study of definite noun expressions in the generative grammar framework over the last decades (Section 2); then, we briefly characterize the languages mentioned with respect to definite nominal constructions (Section 3); finally, we provide a sample of problems directly related to DP acquisition in Spanish and Catalan by learners whose L1 is Russian, Ukrainian, Arabic or Amazigh (Section 4). The article closes with the conclusions (Section 5).

## **2. Nominal structure: the DP analysis**

The study of nominal constructions in the framework of generative grammar has changed substantially since the seminal work of Abney (1987). The Determiner Phrase (DP) hypothesis proposed by Abney gives crucial importance to determiners, which come to be considered as the syntactic head of the whole nominal structure. This line of research has its origins in some parallelisms observed between nominal and sentential constructions in languages like Hungarian and English, and it has led to the establishment of very strong syntactic similarities between the functional and lexical structure of nominal constructions and that of sentences (see Valois 1991, Cinque 1994, Giusti 1997, Aboh 2004 or Svenonius 2004).

Within this framework, grammatical items like the definite article came to be considered as exponents of a functional

nominal category D that would be the akin to functional categories associated to the syntactic structure of sentences like Infl (currently, T), as in the initial proposal by Abney (1987), or C, as originally postulated by Szabolcsi (1987, 1994). Regardless of the precise correspondence with one or other of these categories, all works on this agree in considering that the phrase headed by D forms the extended projection of the noun in the same way that the heads T and C are extended projections of the verb. In the following representations we include the categories *vP* and *nP* as the first functional extension that would contain the generation of the lexical head and its arguments (see Adger 2003 and references therein) but we do not specify the series of functional categories proposed between D and *nP* (see Ritter 1991 or Laenzlinger 2005 for instance).

- (1) a. [<sub>CP</sub> C [<sub>TP</sub> T [<sub>vP</sub> v [<sub>VP</sub> V ] ] ] ]  
 b. [<sub>DP</sub> ..... [<sub>nP</sub> n [<sub>NP</sub> N ] ] ]

One of the first questions raised by the DP hypothesis concerns what is the structure of nominal projections in languages that lack lexical items such as the definite article (i.e. the element that typically occupies the head D). Two possibilities arise a priori: (i) in such languages there is no DP projection (see Bošković 2005), or (ii) the DP projection is present in the syntactic structure in all languages, but whereas in some of them D is explicitly realized by a lexical item, in others it remains empty (Longobardi 1994, Bernstein 2001). We exemplify the two types by means of English and Serbo-Croatian (examples from Bošković 2008):

- (2) English  
 a. The stone *broke the window*.  
 Serbo-Croatian  
 b. Kamen *je razbio prozor*.  
 stone is brokenwindow  
 'The stone broke the window.'



The idea that the head D is syntactically present, but phonologically empty, is consistent with Longobardi (1994)'s influential analysis. This author considers that the DP projection is fundamental for the nominal expression to function as an argument. In his analysis, the presence of the DP ensures the nature of argument and the head D can be filled by a determiner (a head D) or by an element shifted to D after an explicit syntactic movement, or it can remain empty in the syntactic representation awaiting a shift in the Logical Form.<sup>1</sup> The relevance of the DP hypothesis lies in the fact that it offers an interesting framework for syntactic comparative studies, whether approached from a macroparametric perspective (Bošković's NP/DP parameter) or a microparametric one (for example, the use or not of the definite article before the proper name in close Romance variants).

Within the particular field of Romance languages, the DP hypothesis has aroused special interest for diachronic and acquisition analyses. From the point of view of diachronic evolution, the question focuses on the fact that in all Romance languages the existence of a definite article is general, a grammatical piece which did not exist in Latin, whose origin is to be found in the Latin pronominal forms *ille* and *ipse*. Examples such as the following are often taken as indications of an incipient use as a definite article of these forms in Late Latin (example, glosses and translation taken from Ledgeway 2012: 90):

- (3) Latin (*Peregrinatio Aeth* 1.1-2.1)  
 montes illi inter quos ibamus, aperiebat et  
 mounts.N these.A among which.A went.1PL opened and  
*faciebant uallem infinitam [...]* Uallis autem ipsa ingens  
 made valley endless.A valley.N but self.N huge.N

---

<sup>1</sup> Longobardi applies this analysis to proper names and bare NPs in English and Romance languages, which are expressions that act as arguments, but are not introduced by any determiner.

*est ualde.*

is truly

‘th(os)e mountains, through which we were journeying, opened and formed an endless valley. [...] The (= aforementioned) valley is indeed truly huge.’

The definite articles of Romance languages have their origin in two Latin pronouns: *ille*, in most languages (Spanish and Catalan *el*, Italian *il*, French *le*, etc.) and *ipse* in the case of Sardinian *su* and Catalan *es* (see Ledgeway 2012, Ledgeway and Maiden 2016). Originally, the former was a distal demonstrative pronoun and the latter an emphatic pronoun, but both were employed too to establish anaphoric relations and to refer to entities familiar to the interlocutors (though not generic, unique or abstract referents, in contrast to definite articles current usage).<sup>2</sup> Bearing this in mind, we can ask ourselves what is the role of pronominal elements like these in the emergence of a nominal functional category such as determiners. Generativist analyses have formalized this evolution through a process of grammaticalization and reanalysis of the Latin pronoun as the head D, whether it was a head of a lower projection in the nominal structure (4a) or a constituent in the specifier of DP (4b) (see Batllori and Roca 2000 and Giusti 2001, respectively):

- (4) a. [DP [D' [D ] [DemP [Dem' [Dem ille] ... ] > [DP [ D ille] ...  
 b. [DP [ DemPille ] [ D' [D ] ... ] > [ [ DP [ D' [ D ille] ...

From the point of view of second language acquisition, the question is also interesting because since the 1990s generativist studies take functional categories (i.e. their characteristics and the formal features which they are associated with) as the locus for parametric variation and, in consequence, studies on L1 and L2 acquisition granted a crucial role to the develop-

<sup>2</sup> The nature of *ille* and *ipse* as a kind of article in Late Latin is controversial (see Ledgeway 2012: § 4.2.2.1 and references cited).

ment of these categories (see White 2003, Licerias *et al.* 2008 or Meisel 2011, among many others).<sup>3</sup> In the particular case of L2 acquisition the approaches were opened to identify problems related both to the consecutive stages of development of this item and its grammatical feature and to the influence of the L1, especially when it presented notable differences and involved a different parametric choice. In the particular case of definite determiners, the following situations arise: (i) in the L2 the head D is overtly realized by means of a wide range of determiners, but the L1 lacks elements of the D-type; (ii) both L1 and L2 have definite determiners that occupy the head D, but they express different grammatical features in each language; (iii) the L2, but not the L1 lacks definite determiners. The first one is the usual situation we found in the acquisition of Romance languages as L2 with learners whose L1 is a Slavic or Amazigh language; the second one is found in cases of acquisition as L2 of Romance definite article, which are inflected for gender and number, when L1 articles are not inflected (or are inflected in a different set of features), as in Arabic.

The literature on DPs over the past few decades has revolved around the existence (or not) of a highly articulated structure with a long series of functional projections and around the grammatical features and lexical items with which each one is associated. We will not pursue this discussion and we will limit ourselves to assuming (i) that DP is the highest functional nominal projection, (ii) that the definite article is realized in the head D and that there is a functional space below it (labelled “D2” or “lower D”) where certain definite determiners (or elements akin to them can operate in syntax. Following Bernstein in his dialectal and diachronic comparison

---

<sup>3</sup> We are using the acronym L1 as a synonym for mother tongue and L2 to refer to any language that has been learned after having acquired the L1. We do not enter here in further specifications such as L3 (or Ln), which will be pertinent for several groups of individuals in our study at later developments of our research.





(8a) and in phrases that denote unspecific referents (8b), and it is interpreted with possessive value in certain relations of inalienable possession (8c):<sup>5</sup>

(8) Spanish

- a. Los            *leones son peligrosos.*  
 DEF.M.PL lions are dangerous  
 'Lions are dangerous.'
- b. *Hablaré solo con los que suspendan*  
 talk.FUT.1SG only with DEF.M.PL that fail.SUBJ.3PL  
*el examen*  
 DEF.M.SG exam
- c. *Me rompí la mano.*  
 me broke.1SGDEF.F.SG hand  
 'I broke my hand.'

In Spanish, demonstratives distinguish three degrees of deixis: proximity to the speaker (*este*), proximity to the hearer or intermediate distance (*ese*), and distance from both interlocutors (*aquel*). In Catalan, there are two different systems: a binary one that distinguishes only the degrees of proximity (*aquest*) and distance (*aquell*) with respect to the deictic center, and a ternary one with the same distinctions as in Spanish (*este*, *eixe* and *aquell*). The binary system is the most extended and the ternary system is characteristic of some western variants like, for instance, Valencian.

Demonstratives reproduce the same inflection and agreement patterns of the definite article. The main syntactic differences between demonstratives and the definite article are that demonstratives can appear alone (9) or in postnominal position (10):

---

<sup>5</sup> In Catalan, human proper nouns are usually introduced by an article. In some dialectal varieties a specific determiner (the so-called personal article) *en/na* is used; in others, the determiner is the definite article *el/la* (see Brucart 2002: § 7.3.4; IEC 2016: §16.3.1.2). In Spanish, the use of the definite article before a human proper name is also possible, but it is less general and it is subject to social and dialectal variation (see RAE-ASALE 2009: §12.7).

- (9) Catalan
- a. *Agafaré*      *aquell*      *llibre*.  
 take.FUT.1SG that.M.SG book  
 'I will take that book.'
- b. *Agafaré*      *aquell*.  
 take.FUT.1SG that.M.SG  
 'I will take that one.'
- (10) Spanish
- a. *El*              *llibre* *aquel*  
 DEF.M.SG book that.M.SG  
 'That book'
- b. *El*              *noi* *aquest*  
 DEF.M.SG boy this.M.SG  
 'This boy'

Postnominal demonstratives require the presence of the definite article: *\*llibre aquel*, *\*noi aquest*. These constructions have been considered as evidence for a complex nominal structure with two levels of determination along the lines of (5) (see Roca 1997 or, on different grounds, Zamparelli 2000) or with several functional categories that may host the demonstrative (see Giusti 1997, Brugè 2002).

Possessives are different grammatical elements in Catalan and in Spanish. Only Spanish has a possessive that behaves as a determiner, that is, it is prenominal, it converts the nominal expression into definite and allow it to act as an argument, and it is in complementary distribution with the definite article and the demonstrative:<sup>6</sup>

- (11) Spanish
- a. *mi*              *llibre*.  
 POS.1SG book  
 'my book.'

---

<sup>6</sup> The prenominal possessive co-occurred with the definite article and with indefinites in older stages of the language. The co-occurrence with the demonstrative (*esta su casa* 'lit. this your house') is maintained in some western European Spanish dialects or with an archaic flavor (see. RAE/ASALE 2009: §17.4z).

b. *	el	mi	libro/*	mi	el	libro
	DEF.M.SG	POS.1SG	book	POS.1SG	DEF.M.SG	book
c. *	este	mi	libro/*	mi	este	libro
	this.M.SG	POS.1SG	book	POS.1SG	this.M.SG	book

Spanish possessive determiners have person features that agree with the possessor and differ from other determiners in that they do not agree in gender with the noun, unless they are first or second person plural:

(12) Spanish

a.	mi	libro/	casa
	POS.1SG	book.M	house.F
	'my book/house'		
b.	nuestro	libro/	nuestra casa
	POS.1SG.M	book.M	POS.1SG.F house.F
	'Our book'/ 'Our house'		

The monosyllabic prenominal possessives *mi*, *tu* and *su* are like French prenominal possessives and they are clitic forms (see Escandell 1999).

The Catalan possessive paradigm has prenominal forms like those of Spanish and French, but their use is reduced to express certain family ties (*mon pare* 'my father', *ta germana* 'your sister') and, less systematically, with elements that maintain a particularly close relationship with the possessor (*mon poble* 'my village').<sup>7</sup> The possessive generally used in Catalan is

<sup>7</sup> Kinship terms present particular uses in several languages. Thus, in Catalan (and in Spanish too) the presence of the definite article might be enough to interpret the possession relationship when the context is clear (ia), in certain Spanish varieties names like *padre* or *madre* may appear without any determiner (ib) (see RAE-ASALE 2009: 18.7m), and, as an anonymous reviewer noticed to us, in Polish the possessive is often dropped with similar names and situations (ic):

- (i) a. *La Clara ha deixat els fills amb l'àvia.* (Catalan)  
 DEF.F Clara has left DEF.PL sons with DEF-grandma  
 'Clara left her sons with their grandmother.'
- b. Spanish  
*Hablé con padre ayer.*  
 talked.1SG with father yesterday



a stressed form that reproduces the person features of the possessor and agrees in gender and number with the noun that denotes what is possessed. This possessive can be prenominal or postnominal:

## (13) Catalan

a. *el*            *meu*        *llibre/ el llibre* *meu*.

DEF.M.SG POS.1SG book.M

'My book'

b. *la*            *vostra*      *casa/ la casa* *vostra*.

DEF.F.SG POS.1PL.F house.M

'Your (pl) house'

In this sense, Catalan is similar to Romance languages such as Italian (*il mio libro* 'my book') or Portuguese (*o meu livro* 'my book'). The construction with the possessive in prenominal position is usually interpreted as definite, but the definite article must head the construction.<sup>8</sup> This behavior shows that, unlike the case of Spanish (or French and English), this Catalan possessive does not fulfill the syntactic and semantic roles (definiteness, argumenthood, etc.) associated with D in this language.<sup>9</sup>

---

'I talked to my/our father yesterday.'

## c. Polish

*Rozmawiałem z tatą.*

spoke.1SG.M with father.1SG

'I spoke to my father.'

<sup>8</sup> In some varieties, it can also be headed by a demonstrative (*aquest teu amic* 'lit. this your friend') or by an indefinite (*un meu amic* 'lit. 'a mine friend'). In the latter case, which is characteristic of certain areas of Central and Eastern Catalan, the indefinite is interpreted as specific (see Brucart 2002: § 7.5.2.1, IEC (2016: §16.5.1d)).

<sup>9</sup> Only prenominal possessives that are incompatible with the definite article are associated to definiteness and to the D head. This is the case of monosyllabic possessives (like Sp. *mi*, Fr. *mon* or Cat. *mon*) and forms like *nuestro* 'our' or *vuestro* 'your' in Spanish. These forms can be prenominal or postnominal (*nuestro libro* 'our book', *este libro vuestro* 'this book of yours'), but only if they are prenominal can they be related to D (like monosyllabic possessives), are they incompatible with the definite article (*\*el vuestro libro*), and is the expression interpreted as definite (*\*un vuestro libro* vs. *un libro vuestro* 'a book of yours').

### 3.2. Definite nominal constructions in Slavic languages

Most Slavic languages lack definite articles,<sup>10</sup> but not demonstratives or possessives, which, according to Bošković, are elements closer to adjectives that are inflected in gender, number and case, and agree in these grammatical features with the noun they modify. The definite interpretation of a nominal phrase in, for instance, Russian in (14) is deduced from the discourse or pragmatic context and it correlates with the position it occupies in the sentence. Sentence-initial positions are often linked to definite interpretations:<sup>11</sup>

(14) Russian

a. *Ánna citáet knígu.*

Anna.N.SG reads book.ACC.SG

'Anna reads a book.'

b. *Kníga bylá napisana na ispánskom.*

book.ACC.SG was.F.written.NOM.F.SG in Spanish.PREP.SG

'The book was written in Spanish.'

The forms *knígu* and *kníga* have no formal marking that reveals their definite or indefinite character, but context (in this case the position in the sentence) provides this interpretation.

Slavic demonstratives distinguish two degrees of deixis, like the binary system of Catalan and most of the languages collected in Diessel (2013): *cěj* 'this' and *tój* 'that' in Ukrainian; *étot* 'this' and *tot* 'that' in Russian. The demonstrative heads the nominal construction, agrees in gender, number and case with the noun, and, as shown in (15b), may appear without an explicit noun:

<sup>10</sup> As indicated by Bošković, Bulgarian and Macedonian are exceptions.

<sup>11</sup> Russian and Ukrainian examples are taken, respectively, from Chernova and Roca (2010) and Roca (2005) unless otherwise indicated.

## (15) Ukrainian

- a. Ta knýžka ne mojá, a johó.  
 this.SG book.NOM.SG NEG POS.NOM.3SG.F but POS.NOM1SG.M  
 'That book is not mine, but his.'
- b. Ta ne mojá, a johó.  
 this.SG NEG POS.NOM.3SG.F but POS.NOM1SG.M  
 'That one is not mine, but his.'

Possessives express the person features of the possessor, have an adjectival character and, generally, precede the noun, with which they agree in the three grammatical specifications:<sup>12</sup>

## (16) Russian

- a. Ja vzjál tvoju súmku.  
 I took.m POS.ACC.2SG.F bag.ACC.3SG.F  
 'I took your bag.'
- b. Onánáša učitel'nica.  
 she POS.NOM.1PL.F teacher  
 'She is sour teacher.'

Reflexive possessives keep the gender, number and case agreement with the noun, but they do not reproduce the person features of the possessor. In Ukrainian, Russian or Polish the forms *svij*, *svoj* and *swój* are used with antecedents of any of the three grammatical persons:<sup>13</sup>

<sup>12</sup> Third person possessives show a different behavior. They do not agree with the noun and express the gender of the possessor:

- (i) Russian
- a. jegó sestrá  
 he.GEN.M.SG sister.N.F.SG  
 'his sister'
- b. jejó knígi  
 she.GEN.F.SG book(F.).N.PL  
 'her books'

These forms are, as indicated in the glosses, genitive personal pronouns, rather than possessives. They cannot be linked to a D head because the case feature attributed to the whole nominal construction is nominative, as shows the case inflection of the nouns *sestrá* and *knígi*.

<sup>13</sup> We thank an anonymous reviewer for pointing this to us.

## (17) Ukrainian

- a. *Ja vz'av svojú val'izku.*  
 I took.M REF.POS.ACC.SG.F suitcase.ACC.SG.F  
 'I took my suitcase.'

## Russian

- b. *Ty pómniš' svojú škólu.*  
 you remember.2SG REF.POS.ACC.SG.F school.ACC.SG.F  
 'You remember your school.'

## Polish

- c. *Janek czyta swoją książkę.*  
 Janek reads REF.POS.ACC.SG.F book.ACC.SG.F  
 'Janek reads his own book.'

There is no formal difference between possessives that appear in nominal constructions and those that appear as predicates. This suggests that the possessive does not contain grammatical features or properties as the ones of determiners. In this sense, Slavic possessives are similar to those of Catalan (and other Romance languages) but different to those of English or Spanish:

## (18) Russian

- a. *Étot karandáš moj.*  
 This.NOM.SG.M pencil.NOM.SG.M POS.1SG.NOM.SG.M  
 'This pencil is mine.'
- b. *moj karandáš*  
 POS.1SG.NOM.SG.M pencil.NOM.SG.M  
 'My pencil'

## (19) Catalan

- a. *Aquest llapis és meu.*  
 This.SG.M pencil.SG.M is POS.1SG.M  
 'This pencil is mine.'
- b. *El meu llapis*  
 the POS.1SG.M pencil.SG.M  
 'My pencil'

## (20) English

- a. *This pencil is mine / \*my.*  
 b. *My pencil*

- (21) Spanish
- a. *Este lápiz es mío/\*mi.*  
 This.SG.M pencil.SG.M is POS.1SG.M  
 'This pencil is mine.'
- b. *Mi lápiz.*  
 POS.1SG.M pencil.SG.M  
 'My pencil'

In conclusion, the preceding data suggest that neither demonstratives nor possessives would unambiguously correspond to the realization of the head D of the highest DP projection: possessives do not have determiner-like properties at all; demonstratives coincide with Romance definite determiners in the definite interpretation, in the initial position and in constructions without an overt noun. These properties could be related to a lower position different to the one that occupies the definite article.

### 3.3. Definite nominal constructions in Egyptian Arabic<sup>14</sup>

In contrast with Slavic languages, Egyptian Arabic (also known as Ameya) has a definite article that can be considered as the head of the DP. However, this element is not the only way to mark that the nominal expression is interpreted as definite. In this language definiteness may be expressed through different ways: by means of the definite article *al-* (22a), a genitive complement (22b), a possessive affix (22c), or a demonstrative that follows the noun preceded by the definite article (22d).<sup>15</sup>

<sup>14</sup> Arabic language is a set of varieties that present notable differences among them. Here we focus on describing one of these varieties: Egyptian Arabic or Ameya.

<sup>15</sup> We simplified the transliteration of Arabic vowels in three (<a>, <i>, <u>), as dictated by modern standard Arabic, though in the spoken language there are the allophones [e] and [o]. We follow Bezos (2006) transliteration system and we adapt the pronunciation of words according to Ameya except in cases where we give standard Arabic examples.

- (22) Ameya
- a. *al-walād*  
DEF-boy  
'The boy'
  - b. *tālib al-ġam'ah*  
student def-university  
'the student of the university.'
  - c. *kitāb-hu*  
book-POS.3SG.M  
'his book'
  - d. *al-kitāb dā*  
DEF-book this.SG.M  
'this book.'

Definiteness grammatical markers are important in Arabic DPs. The following examples differ only in the presence of *al-* adjoined to the adjective *tāyib* 'good'. In definite DPs, definiteness spreads to the adjective and the definite article must appear with both the noun and the adjective (23a); if the adjective does not bear this definiteness marker, the sequence is interpreted as a copulative sentence (23b):

- (23) Ameya
- a. *al-walād al-tāyib*  
DEF-boy DEF-good  
'The good boy'
  - b. *al-walād tāyib*  
DEF-boy good  
'The boy is good.'

Definiteness spreading through elements inside the DP clearly distinguishes Arabic from Romance languages, where it is expressed only by means of one element in the head D, and from Slavic languages, where it is not expressed by any overt element in D. The Arabic definite article *al-* is proclitic to the noun or to the adjective and it lacks gender and number features (Corriente 1988). Its main function is to set the reference

of the nominal expression; in its absence, the whole nominal expression is interpreted as indefinite:

- (24) Ameya  
*walād tāyib*  
 good boy  
 ‘a good boy’

Thus, Arabic bare NPs are clearly different to Slavic bare NPs: whereas the first ones correlate with indefiniteness, the second ones are potentially ambiguous between a definite and an indefinite interpretation.

In Ameya, the definite article is also used in contexts with non-specific referents phrases interpreted as generics (25) and in certain inalienable possession relationships (26):

- (25) Ameya  
 a. *Al-qahua<sup>h</sup> bita-rfa’ al-duġat.*  
 DEF-coffee brings up DEF-pressure  
 ‘Coffee brings pressure up.’  
 b. *Al-līmūn fākha<sup>h</sup>*  
 DEF-lemon fruit  
 ‘A lemon is a fruit.’  
 c. *Al-kilāb ‘and-hā ārba’ riglin.*  
 DEF-dogs in-POS.3SG.F four.M legs.F  
 ‘Dogs have four legs.’

- (26) Ameya  
*Qal’nā al-baranit*  
 take.1PL.PF DEF-hats.PL  
 ‘We took off our hats.’

In standard Arabic singular nouns preserve the specific reading unless the presence of a predication or an adverb induce the generic interpretation (see Fassi Fehri 2007), but in Ameya sentences like those in (25) are interpreted as generic or as referential simply according to the pragmatic context. The use

of the definite article in generic in Arabic is, then, akin to the one we find in Spanish or Catalan.<sup>16</sup>

Concerning inalienable possession relations, Ameya distinguishes between parts of the body and clothes. As shown in (26), clothes are introduced by means of a DP headed by the definite article (*al-baranit* 'the hat'), as in Spanish (*Nos quitamos los sombreros* 'We took off our hats'), but parts of the body need the possessive (as in English):

- (27) Ameya  
*Mary bita-rfa' yid-hā.*  
 Mary IMPF.3SG.F-raise hand-3SG.F  
 'Mary raises her hand.'

Following Hänninen (2014) we consider, then, that Ameya grammar is sensitive to the nature of the semantic relationship between the object and the possessor and that the highest degree of alienation is encoded through the need to use the possessive.

In Ameya, as in Spanish, it is possible to drop the noun in presence of a modifier:

- (28) Ameya  
 a. *al-kitāb al-ahmar.*  
 DEF-book DEF-red  
 'the red book'  
 b. *al-ahmar*  
 DEF-red  
 'the red one'

In standard Arabic demonstratives express a two-way deictic distinction: proximity and distance from the speaker, as in English, Russian, Ukrainian, or Catalan. However, in Egyptian

---

<sup>16</sup> Arabic contrasts with English and Romance languages in the behavior of indefinites: indefinite nominal phrases can be interpreted as generic in English (*A whale is a mammal*) or in Spanish (*Una ballena es un mamífero*), but in Arabic they are interpreted with an existential reading (see Fassi Fehri 2007: 47).



Arabic the same form of demonstrative is used to express both proximity and distance. In terms of Diessel (2013), this would be a case of no distance contrast system, where the identification of distant referents is made by the pragmatic context, which provides the proximity or distal interpretation:

- (29) Ameya  
*al-raġul* *dà*  
 DEF-man this.M  
 'this/that man'

Demonstratives show variation in gender and number (*da* 'this/that' is masculine singular, *dī* 'this/that' is feminine singular, and *dī* or *dol* 'these/those' is plural) and agree with the noun they are referring to, which can be explicit or not:<sup>17</sup>

- (30) Ameya  
 a. *al-raġul* *dà*  
 DEF-man this.M  
 'this/that man'  
 b. *al-bint* *dī*  
 DEF-girl this.F  
 'this girl'
- (31) Ameya  
*Ajat* *dà*.  
 take.PF that  
 'I took this/that one.'

Differently to the definite article and to possessives (see below), Arabic demonstratives are not clitics. Their usual position in Standard Arabic is prenominal, but, in Ameya they appear in postnominal position (and the definite article must precede the noun, as in Spanish or in Catalan):

---

<sup>17</sup> The plural of non-human nouns are treated as feminine, this means that they will take the demonstrative *dī* (this treatment extends to adjectives). The Standard Arabic number distinction between dual and plural is not preserved in Ameya, where the plural form subsumes the dual.

- (32) Ameya  
*al-bint dī*  
 DEF-girl this.F  
 'this girl'

Possessives are clitic morphemes attached to the right of nouns, pronouns, verbs or prepositions. When combined with a noun, the possessive provides the meaning of possession and turns the nominal construction into definite, as Spanish pre-nominal possessives do. The definiteness content provided by the possessive is consistent with the fact that it cannot co-occur with the definite article:

- (33) Ameya  
 a. *bayt-y*  
 house-POS.1SG  
 'my house'  
 b. *\*āl-bayt-y*  
 def-house-POS.1SG

The possessive determiner only expresses the person and number features of the possessor; it does not show any kind of agreement with the noun (the possessive). In this sense, it is closer to English possessive forms than to Slavic or Romance forms, where agreement with the noun is expressed.

The possessive can co-occur with a demonstrative, which follows the noun with the possessive:

- (34) Ameya  
*kitāb-nā dā.*  
 book-POS.1PL this.M  
 'This book of ours'

In sum, Ameya has a definiteness marker that can occupy the head D that introduces the nominal construction. Demonstratives and possessives do not display such a behavior and they

should be considered as elements that contribute to the definite interpretation but appear in a lower position.

### 3.4. Definite nominal constructions in Amazigh<sup>18</sup>

Amazigh is a language that, like Slavic languages or Latin but unlike Romance languages or Arabic, lacks definite articles. In Amazigh the nominal expression is interpreted as definite or as indefinite according to its function in the discourse (example taken from Quitout, 1997 [adapted by Lamuela 2002]):

- (35) Amazigh  
*Tedew* ṭerijat.<sup>19</sup>  
 fly.3SG.PF butterfly  
 ‘A/the butterfly is flying away.’

This is the general situation in most varieties of Amazigh, including those with a larger number of speakers, but this idea is not fully accepted among all Amazigh researchers. Some authors claim that there existed some elements prefixed to nouns (*a-* for the singular, *i-* for the plural) that formerly conveyed the value of a definite article. Traces of this definiteness marker would remain in some varieties. Vycichl (1989) gives examples from Amazigh of Djebel Nefusa, among others, in which the definite/indefinite distinction appears. In the following example, the definiteness marker appears attached to the adjective in a parallel way as the Arabic definite article is enclitic to the adjective that modifies a definite noun (examples from Vycichl, 1989):

---

<sup>18</sup> Amazigh presents a remarkable dialectal variety. Even so, researchers such as Chaker (1995) or Múrcia (2015, 2021) affirm, following both linguistic and sociolinguistic criteria, that it is a single language. With this in mind, in this paper we describe the functioning of determiners in Amazigh as a whole, focusing on specific varieties when considered necessary.

<sup>19</sup> Amazigh is a language in the process of standardization, so there is variation in the encoding of the language depending on the author. In this work we keep the spelling used in the sources consulted unless otherwise indicated.

## (36) Djebel Nefusa Amazigh

- a. *bucīl* aməckân  
 boy DEF.little  
 ‘the little boy’
- b. *bucīl* məccək  
 boy little  
 ‘a little boy’

Amazigh demonstratives distinguish between proximity and remoteness in relation to the first person. This binary distinction turns into a ternary one in some varieties where another demonstrative is used to refer to proximity with respect to the second person. Demonstratives are morphemes that are suffixed to the noun and that do not show any gender or number agreement with it. In the following examples the suffixes *-a* and *-nn* indicate proximity (first person) and remoteness (second and third person), respectively (examples adapted from Aghmiri 2014: 22):<sup>20</sup>

## (37) Amazigh

- a. *Argaz-a walu yar-s taddart.*  
 man-this nothing in-PR3SG house  
 ‘This man has no home.’
- b. *Argaz-inn walu yar-s taddart*  
 man-that nothing in-PR3SG house  
 ‘That man has no home.’

In the varieties of Amazigh that express proximity to the second person, like, for instance, Tachelhit, the demonstrative *-a* serves only to indicate remoteness and the suffix *-nna* is used to refer to the second person (Naït-Zerrad 2011): *argaz* ‘man’ vs. *argaz-nna* ‘this man (who is where you are)’.

---

<sup>20</sup> In many dialects the first person demonstrative is *-a* (preceded by [y] if the noun ends in *-a*), but in others it may be *-u* and in some others, as in Tachelhit, it takes the form *-ad* (see Murcia 2021, Naït Zerrad 2011). We made a spelling change:  $\dot{g} > \gamma$  (also in the Aghmiri’s examples that follow).

In addition, there are also demonstratives that have an anaphoric function. This is the case of the suffixes *-(e)nni* in Riffian, *-lli* in Tachelhit, and *-nni* in Kabyle (see Sarrionandía 1905, Naït-Zerrad 2011):

- (38) Riffian Amazigh  
 a. *taddart-nni*  
    house-that  
    ‘that house in question’  
 Tachelhit Amazigh  
 b. *afrux-lli*  
    boy-that  
    ‘that boy in question’  
 Kabyle Amazigh  
 c. *arrac-nni*  
    children-that  
    ‘those children in question’

The anaphoric demonstrative pronoun is also used when the noun is not expressed lexically (see Lamuela 2002).<sup>21</sup> Then, a (non-affix) full form is used preceding the modifier of the empty noun (example from Sarrionandía 1905: 371):

- (39) Riffian Amazigh  
*Necc xsey            wenni        amezyan.*  
 I        want.1SG.PF    that.M.SG    little  
 ‘I want the little one.’

With the exception of 1<sup>st</sup> person singular possessive, which is *inu*, in Amazigh possession is expressed through a possessive constituent that combines the prepositional suffixed pronoun, interpreted as the possessor, with the genitive preposition *n* ‘of’, which can undergo some phonetic modification such as the

---

<sup>21</sup> The demonstratives we have described so far can be attached to other elements, such as pronouns which, as we have seen above, express proximity to the 1st person (*ta d yelli-s n ujellid* ‘this is the king’s daughter’) or to the second and third person (*tin d taddart-inu* ‘that is our house’) or have anaphoric function (*ayenni war ihli* ‘this does not work’).

tension of the consonant *n*, as shown in (40). This combination suffixed to the noun and its presence leads to the definite interpretation of the nominal phrase (examples from Aghmiri 2014: 30):<sup>22</sup>

(40) Amazigh

- a. *taddart-inu*            *taddart-nns*  
 house-POS.1SG        house-POS.3SG  
 ‘my house’            ‘his/her house’
- b. *Tiḍtawin-inu nnumnt akd tfuct cwait cwait.*  
 eyes-POS.1SG get-used-to.3SG with sun gradually  
 ‘My eyes get used to the sun gradually.’

The possessive construction is subject to differences among varieties. In Kabyle, for instance, there is no genitive preposition and the pronominal element preceded by an *-i* is attached to the noun (examples from Kossmann 2012: 75):

(41) Kabyle Amazigh

- aqcic-is*  
 boy-POS.3SG  
 ‘his/her boy’

Finally, possessive constructions with certain nouns, such as kinship terms, involve direct affixation of the possessive to the noun with the particularity that it is preceded by *t* when the possessor is plural (example (42c) from Aghmiri 2014: 27):<sup>23</sup>

<sup>22</sup> When the noun is omitted, the preposition and the pronoun are postponed to the anaphoric demonstrative (Sarrionandia, 1905):

(i) Ncc xsy            *tinni-nkmt*,    war xsy            *tinni-nsnt*.  
 I    want.1SG this-POS2PL.F. not want.1SG. this-POS3PL.F  
 ‘I want yours, not theirs.’

<sup>23</sup> The terms *mmi* ‘son’ and *ylli* ‘daughter’ carry the suffixed pronoun indicating possessor, even when expressed by another N (example (ia) from Lamuela 2002: 50; example (ib) from Kossmann 2012: 76):

(i) Amazigh  
 a. *ylli-s*                            *n*    *wuma*  
 daughter-PR3SG of    brother  
 ‘my brother’s daughter’

- (42) Amazigh
- a. *baba-k*  
 father-POS.2SG.M  
 ‘your father’
- b. *baba-tsn*  
 father-POS.3PL.M  
 ‘their father’
- c. *baba-s*                      *issxdm*                      *Xuan.*  
 Father-POS.2SG.M    make-work.3SG.M    Juan  
 ‘His father made Juan work.’

The different behavior with particular kinds of nouns reminds of the choice of monosyllabic prenominal possessives in Catalan. In comparison with the preceding languages, all the grammatical items (possessives and demonstratives) inducing the definite interpretation in Amazigh appear postnominally. This would indicate that there is no specific determiner for the head D and that, probably, the elements contributing to fix the definite interpretation occupy a lower position in the structure.

### 3.5. Summary

We summarize the main grammatical properties of possessives, demonstratives and definite articles in the six languages in Table 1.<sup>24</sup>

- 
- b. *mmi-s*                      *n Ffaya*  
 son-PR3SG    of Mustapha  
 ‘Mustapha’s son’

<sup>24</sup> The label “Determiner-like” holds for the ability to provide a definite interpretation, head the nominal construction and function as an argument.

**Table 1**  
Definite articles, demonstratives and possessives

	Definite article		Demonstrative			Possessive		
	Position	Agreement with N	D-like	Position	Agreement with N	D-like	Position	Agreement with N
Sp	Pre-N Initial	Gender and number	Yes	Pre-N or post-N Initial	Gender and number	Yes	Pre-N	Number
Cat	Pre-N Initial	Gender, number	Yes	Pre-N/post-N Initial	Gender and number	No	Pre-N or post-N	Gender and number
Rus / Ukr	∅		Yes	Pre-N (generally) Initial	Gender, number and case	No	Pre-N or post-N	Gender, number, case (not with 3rd poss.)
Arab	Pre-N + pre-Adj Initial	No	No	Post-N	Gender and number	No	Post-N (suffixed)	No
Amaz	∅		No	Post-N	Gender and number (only if used as a pronoun)	No	Post-N (suffixed form)	No

The comparison among these languages shows relevant differences with respect to the way in which definite nominal expressions are built in each language. Only in Spanish definiteness is systematically encoded by means of grammatical elements located in the DP projection. Catalan, as well as other Romance languages, also places some elements (demonstratives and the definite article) in this projection, but it locates possessives in a lower position in the nominal functional projection. In Egyptian Arabic the initial D-head position is filled only by the definite article. Slavic languages do not have grammatical elements (i.e. definite articles) that appear in D, but demonstratives show a similar behavior in the sense that they are prenominal, introduce the nominal construction and can appear alone as arguments. Finally, Amazigh does not



seem to have any grammatical element that could be clearly assimilated to the DP projection.

A syntactic structure like the one in (5) hosts elements akin to determiners in the functional structural space below DP. Elements such as demonstratives or possessives, which tend to facilitate the definite interpretation of the nominal construction, might be located in this area. Then, the comparison among these languages can be conceived in terms of the properties of the functional categories in the higher area of the nominal projection. This means, for instance, that Slavic (or Arabic or Amazigh) demonstratives could be considered as determiner-like elements related to a lower D, whereas Romance demonstratives are generally related to the higher D. We can give the following tentative structures for each language:<sup>25</sup>

- (43) a. [DP [D' [D Dem ] [D<sub>2P</sub> [D<sub>2'</sub> [D<sub>2</sub> ] ... [NP ... ] ] ] ] ] *Romance*  
       [DP [D' [D Def Art ] [D<sub>2P</sub> [D<sub>2'</sub> [D<sub>2</sub> Dem ] ... [NP ... ] ] ] ] ]  
       b. [DP [D' [D ] [D<sub>2P</sub> [D<sub>2'</sub> [D<sub>2</sub> Dem ] ... [NP ... ] ] ] ] ] *Slavic*  
       c. [DP [D' [D Def Art ] [D<sub>2P</sub> [D<sub>2'</sub> [D<sub>2</sub> Dem ] ... [NP ... ] ] ] ] ] *Arabic*  
       d. [DP [D' [D ] [D<sub>2P</sub> [D<sub>2'</sub> [D<sub>2</sub> Dem ] ... [NP ... ] ] ] ] ] *Amazigh*

Then, the syntactic properties of the kind of elements related to the expression of definiteness in the higher area (D<sub>2</sub> and D<sub>1</sub>) of the structure appear as very pertinent for the development of the category D. The study of L2 acquisition of definite DPs may shed some light on this issue.

---

<sup>25</sup> We do not specify the syntactic (head and/or phrasal) movement operations that yield the surface linear order and that involve the specifier of DP<sub>2</sub> as a hosting position when the noun precedes the demonstrative ([DP<sub>1</sub> *el* [DP<sub>2</sub> [NP *libro*] [D<sub>2</sub> *este* ... *ti* ... ] ]], for instance) as well as other specifiers in the case of possessives (see Roca 1997, Giusti 1997, Escandell 1999 or Bernstein et al. 2019). A detailed application of this analysis to every language goes beyond the scope of this paper.

#### **4. Acquisition of definite nominal expressions in Spanish and Catalan as second languages**

In this section we present a sample of errors produced by learners of Spanish and Catalan as L2 with Russian, Ukrainian, Egyptian Arabic or Amazigh as L1. The data we analyzed came from acceptability tests, production tasks and semi-structured interviews. The participants were all adults and they started studying Spanish or Catalan after adolescence.<sup>26</sup> This first approach, which consists only of a presentation of some problems detected from a qualitative point of view, is to be continued in future work by ensuring the uniformity of data across all groups of speakers and by incorporating a quantitative analysis of errors.

The use of the definite article and gender agreement are two of the biggest difficulties detected in the acquisition of nominal constructions in Spanish or Catalan as L2. The fact that three of the languages we have described (Russian, Ukrainian and Amazigh) lack definite articles, the canonical realization of the functional head D, suggests that the acquisition of this functional category may be particularly problematic for learners with any of these languages as L1. By contrast, Arabic has a morphologically invariant definite article, so learners with this L1 would be expected to show less difficulties.

Our preliminary analysis shows that this expectation is met, but some with nuances. First of all, it should be noticed that the use of the definite article is attested from the very first levels in all learners; there are no differences depending on the properties of the L1, in this sense. Similarly, errors related to the misuse of bare noun phrases are found with all learners but especially among those with Russian, Ukrainian or

---

<sup>26</sup> The population sample differs from each group of languages in terms of the context where they learn Spanish: the L1 Ukrainian, Russian and Amazigh groups did learn Spanish and Catalan in an immersion context while the group of L1 Ameyya did learn Spanish in Egypt at the *Instituto Cervantes*.

Amazigh as L1. The absence of a determiner or a quantifier occurs with both definite and indefinite expressions and even in constructions in which the noun is omitted (46):

## (44) Catalan

- a.  $\emptyset$  < els > homes regalen a  $\emptyset$  < les > mullers (L1 Russian)  
 DEF men give to DEF wives

'The men give to their wives.'

- b.  $\emptyset$  < els > pobres tenien que pagar la terra (L1 Ukrainian)  
 DEF poor had that pay DEF land

'The poor had to pay for the land.'

## Spanish

- a. Cambiar  $\emptyset$  < la > ropa y el pelo. (L1 Amazigh)  
 change DEF clothes and DEF hair

'They change their clothes and their hair.'

- b.  $\emptyset$  < los > plátanos tienen potasio (L1 Ameya)  
 DEF bananas have potassium

'Bananas have potassium.'

## (45) Catalan

- a. *com  $\emptyset$  < un > arbre, però no és  $\emptyset$  < un > arbre,*  
 like INDEF tree but not is INDEF tree

*és un cérvol.* (L1 Russian)

is INDEF deer

'like a tree, but it's not a tree, it's a deer.'

- b. [*-Com era (la joguina)?*] - *Era  $\emptyset$  < un > ós.* (L1 Ukrainian)  
 how was DEF toy was INDEF bear

'-What did the toy look like? - It was a bear.'

## Spanish

- c. [*- Qué leéis?*] -  $\emptyset$  < un > libro. (L1 Ukrainian)  
 what read INDEF book

'-What are you reading? - A book.'

## (46) Catalan

- a. *Per a mi totes són fàcils, però  $\emptyset$  < la > més difícil*  
 for to me all are easy but DEF more difficult  
*és llengua.* (L1 Russian)

is language

'For me they are all easy, but the most difficult one is language.'

- b. *més bo que Ø < la> d' aquí* (L1 Russian)  
 more good than DEF of here  
 'better than here'

Interestingly, learners with Ameya as L1 produce sequences like the one in (44d), which corresponds to generic noun phrases that in Arabic are introduced by the definite article as in Spanish and Catalan. This indicates that factors other than L1, such as the possible influence of English L2 or general factors of the determiners and definiteness acquisition can influence the development of the definite article during L2 acquisition (see Pérez-Leroux et al. 2004, Rosado 2007, Landa-Buil 2010 or Vilosa 2021). It is also worth mentioning that the omission of the definite article is more frequent than that of the indefinite article: in a rough calculation 3 out of 4 errors of absence of determiner in learners with the two Slavic languages as L1 are of the definite article (76 %).

Overgeneration of the definite article also occurs, but it is considerably less common than those of absence: again, in an approximate calculation, 1 out of 5 (18.6 %) errors related to the use of definite determiners in the two Slavic languages correspond to unnecessary use (as opposed to 75.6 % of errors of absence). Some examples:

(47) Catalan

- a. *els llibres del Ø rus.* (L1 Russian)  
 def books of. DEF Russian  
 'the Russian books.'
- b. *(una nena) que tenia un pare i no tenia la <Ø>*  
 INDEF girl that had a father and not had DEF  
*mare.* (L1 Ukrainian)  
 mother

Spanish

- c. *-Cuánto tiempo hace?- Los <Ø> dos años* (L1 Amazigh)  
 how.much time does DEF two years  
 'How long has it been? - Two years.'

The use of demonstratives and possessives shows few deviations from proper usage in Spanish or Catalan. Problems with gender agreement aside (see below), most errors involve syntactic order (48a), absence of the definite article (48b-d), and redundant use in inalienable possession relations (48e):

- (48) Catalan
- a.  $\emptyset$  < les > festes de Tura aquestes. (L1 Amazigh)  
 DEF feast.daysof Tura these  
 'These feast days of Tura.'
- b. para  $\emptyset$  < el > nostre pare (L1 Ukranian)  
 for DEF our father  
 'for our father'
- c. jo amb  $\emptyset$  < el > seu marit. (L1 Russian)  
 I with DEF his/her husband  
 'me and his/her husband.'
- d. Va posar el seu <  $\emptyset$  > cap dins del pot. (L1 Russian)  
 put.3sg DEF his head into of.DEF pot  
 'He put his head inside the pot.'
- Spanish
- e. Me duele mi < la > cabeza. (L1 Ameya)  
 me hurts my DEF head  
 'I have a headache.'

According to the characteristics of their L1, learners are likely to consider that the presence or the demonstrative in (48a) in a possible position in the L2, but under certain circumstances, or the possessive in (48b-d) is enough to turn the nominal expression into definite and, consequently, they dispense with the obligatory initial determiner. In (48e) the inadequate use of the possessive reflects the use in Arabic.

Although our main interest lies in the acquisition of determiners, we will conclude this section with a couple of remarks on problems detected in the realization of gender. The acquisition of gender and gender agreement is one of the main difficulties in the acquisition of Spanish or Catalan as an L2 and

leads to fossilized errors (see Alarcón 2011). Some examples of gender confusion in definite phrases are the following

(49) Catalan

a. *les* <*els*> *dies, els* <*les*> *cançons*  
 DEF.F.PL DEF.M.PL days DEF.M.PL DEF.F.PL songs  
 (L1 Russian)

b. *la* <*el*> *seu pare, les ulls* <*els*>  
 DEF.F.SG DEF.M.SG POS.3SG.M father DEF.F.PL DEF.M.PL  
 eyes  
 (L1 Amazigh)

Spanish

c. *los* <*las*> *ratas, las billetes.*  
 DEF.M.PL DEF.F.PL rats DEF.F.PL DEF.M.PL notes  
 (L1 Ukranian)

d. *el* <*la*> *ciudad, la* <*el*> *coche.*  
 DEF.M.SG DEF.F.SG city DEF.F.SG DEF.M.SG car  
 (L1 Arabic)

Among the reasons for the confusion in the assignment of grammatical gender to nouns we found the arbitrariness of gender in Spanish and Catalan or the lack of a systematic formal correlation with the L1, which can result in positive transfer or negative interference (see, for instance, Sabourin et al 2006 or Vilosa 2021). In relation to the nature of the L1, we have noticed a remarkable difference on learners with the two Slavic languages as L1 and those with Amazigh or Arabic: in the former, gender confusions appear relatively balanced between masculine and feminine (although with a slight preference for the masculine), but in the latter, most of the errors are due to the use of the feminine instead of the masculine form. This difference might indicate that the Indo-European versus non-Indo-European character of the languages involved is somehow influencing the acquisition process.

This difference reflects the difficulty learners have in familiarizing themselves with the use of this determiner and suggests that the acquisition (or development) of the higher func-

tional layer (the DP) of the nominal structure is particularly problematic (for L1 speakers who lack this element).

## **5. Conclusion**

In this article we have compared the nominal constructions of typologically different languages and we have provided a first piece of evidence for the difficulties to acquire the functional projection D by learners with one of these languages as L1. Our preliminary analysis shows, on one hand, that the acquisition of definite nominal expressions is problematic in general and, to some extent, independent of the properties of the L1, and, on the other hand that there are relevant differences among definite determiners (or elements associated with the DP projection in the L2). These differences can be related to difficulties in acquiring a good command of the highest functional field in the nominal structure, where the distinction between several layers of determination can help to identify the problems posed by certain lexical items and to find correlations in the development of the category D in Romance languages.

Our future steps in this research are, firstly, to get a more balanced set of informants and data in order to be in a better position to carry out qualitative and quantitative analyses, and secondly, we will deepen our knowledge of the grammatical characteristics and processes of the different L1 and of the possible influence of other L2. The data and conclusions provided by a study on the acquisition of nominal structure from this comparative perspective are also relevant for heritage languages, where the emergence of a DP projection can also occur as a consequence of language contact.

## References

- Abney, Steven (1987). *The English Noun Phrase in Its Sentential Aspect*. PhD dissertation, MIT.
- Aboh, Enoch O. (2004). "Topic and focus within DP". *Linguistics in the Netherlands* 21: 1-12.
- Adger, David (2003). *Core Syntax: A Minimalist Approach*. Oxford: Oxford University Press.
- Alarcón, I. V. (2011). "Spanish gender agreement under complete and incomplete acquisition: Early and late bilinguals' linguistic behavior within the noun phrase". *Bilingualism: Language and Cognition* 14: 332-350.
- Aghmiri, Rachid (2014). *Los usos de los 10 verbos más frecuentes del español comparados con el amazigh*. Master's dissertation, Universitat de Girona.
- Basset, André (1969). La langue berbère. *Handbook of African Languages*. Part I. London: Routledge.
- Batllo, Montserrat, Francesc Roca (2000). "The value of definite determiners from Old Spanish to Modern Spanish". In: Susan Pintzuk et al. (eds.) *Diachronic Syntax: Models And Mechanisms*. Oxford: Oxford University Press, 241-254.
- Bernstein, Judy (2001). "The DP hypothesis". In: M. Baltin, C. Collins (eds.). *The Handbook of Contemporary Syntactic Theory*. Oxford: Blackwell, 537-561.
- Bernstein, Judy, Francisco Ordóñez, Francesc Roca (2019). "On the emergence of personal articles in the history of Catalan". In: M. Bouzouita, A. Breitbarth, L. Danckaert, E. Witzhausen (eds.). *Cycles in Language Changes*. New York – Oxford: Oxford University Press, 88-108.
- Bernstein, Judy, Francisco Ordóñez, Francesc Roca (2021). "Emphatic elements and the development of definite articles: Evidence for a layered DP in early Romance". *Journal of Historical Syntax* 5: 1-32.
- Bezós, Javier (2006). "Sistema de transliteración". *Panacea* VII/23: 149.
- Bošković, Željko (2008). "What will you have, DP or NP?". In Emily Elfner, Martin Walkow (eds.). *Proceedings of NELS 37*. Cambridge, MA: Graduate Linguistic Student Association, 101-114.



- Bošković, Željko (2012). "On NPs and clause". In: G. Grewendorf, T. E. Zimmermann (eds.). *Discourse and Grammar: From Sentence Types to Lexical Categories*. Boston – Berlin: Mouton De Gruyter, 179-242.
- Bošković, Željko (2005). "On the locality of left branch extraction and the structure of NP". *Studia Linguistica* 59/1: 1-45.
- Brucart, Josep Maria (2002). "Els determinants". In: J. Solà (ed.). *Gramàtica del català contemporani*. Barcelona: Empúries, 1435-1516.
- Brugè, Laura (2002). "The positions of demonstratives in the extended nominal projection". In: G. Cinque (ed.). *Functional Structure in DP and IP*. Oxford: Oxford University Press, 15-53.
- Chaker, Salem. (1995). "Dialecte". *Encyclopedie berbère*. Vol. XV. Aix-en-Provence: Edisud, 2291-2295.
- Chernova, Ekaterina, Francesc Roca (2010). *El rus: Estudi comparatiu entre la gramàtica del català i la del rus*. Barcelona: Fasolcat.
- Cinque, Guglielmo (1994). "On the evidence for partial N-movement in the Romance DP". In: G. Cinque, J. Koster, J-Y. Pollock, L. Rizzi, R. Zanuttini (eds.). *Paths towards Universal Grammar*. Washington D.C.: Georgetown University Press, 85-110.
- Corriente, Federico (1988). *Gramàtica àrabe*. Barcelona: Herder.
- Diessel, Holger (2013). "Distance contrasts in demonstratives". In: M. S. Dryer, M. Haspelmath (eds.). *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Available at <<http://wals.info/chapter/41>>. Accessed 15.10.2021.
- Escandell, Victoria (1999). "Notas sobre la gramàtica de los posesivos". In: P. Carbonero Cano, M. Casado Velarde, P. Gomez Manzano (eds.). *Lengua y discurso: Estudios en honor del profesor Vidal Lamiquiz*. Madrid: Arco Libros, 265-277.
- Fassi Fehri, Abdelkader (2007). "Bare, generic, mass, and referential Arabic DPs". In: S. Karimi, V. Samiian, W. K. Wilkins (eds.). *Phrasal and Clausal Architecture: Syntactic Derivation and Interpretation*. Amsterdam – Philadelphia: John Benjamins, 40-65.
- Fassi Fehri, Abdelkader (2018). *Constructing Feminine to Mean: Gender, Number, Numeral and Quantifier Extensions in Arabic*. Lanham: Lexington Books.
- Franceschina, Florencia (2001). "Morphological or syntactic deficits in near-native speakers? An assessment of some current proposals". *Second Language Research* 17/3: 213-247.

- Giusti, Giuliana (1997). "The categorial status of determiners". In: L. Haegeman (ed.). *The New Comparative Syntax*. London: Longman, 95-123.
- Giusti, Giuliana (2001). "The birth of a functional category: From Latin *ille* to the Romance article and Personal pronoun". In: G. Cinque, G. Salvi (eds.). *Current Studies in Italian Syntax*. Amsterdam: Elsevier, 159-171.
- Hänninen, Mira (2014). *Posesión inalienable: La enseñanza y los conocimientos lingüísticos de los estudiantes finlandeses de ELE*. Master's dissertation, Helsinki University.
- IEC (2016). *Gramática de la llengua catalana*. Barcelona: Institut d'Estudis Catalans.
- Kossmann, Maarten (2012). "Berber". In: Z. Frajzyngier, E. Shay (eds.). *The Afroasiatic Languages*. Cambridge: Cambridge University Press, 18-101.
- Laenzlinger, Christopher (2005). "French adjective ordering: Perspectives on DP-internal movement types". *Lingua* 115: 645-689.
- Lamuela, Xavier (2002). *El berber: Estudi comparatiu entre la gramàtica del català i la del berber o amazig*. Barcelona: Generalitat de Catalunya, Departament de Benestar i Família.
- Landa-Buil, María (2010). *El sintagma determinante en la interllengua espanyola de parlants de swahili*. PhD dissertation, Universidad Complutense-Instituto Universitario de Investigación Ortega y Gasset.
- Ledgeway, Adam (2012). *From Latin to Romance*. Oxford: Oxford University Press.
- Ledgeway, Adam; Maiden, Martin (2016). *The Oxford Guide to Romance Languages*. Oxford: Oxford University Press.
- Liceras, Juana, Helmut Zobl, Helen Goodluck (eds.) (2008). *The Role of Formal Features in Second Language Acquisition*. London – New York: Routledge.
- Longobardi, Giuseppe (1994). "Reference and proper nouns". *Linguistic Inquiry* 25: 609-665.
- Meisel, Jürgen M. (2011). *First and Second Language Acquisition*. Cambridge: Cambridge University Press.
- Múrcia, Carles. (2015). "Compendi gramatical". In: C. Múrcia, S. Zenia. *Diccionari català-amazic, amazic-català*. Barcelona: Llibres de l'index.

- Múrcia, Carles (2021). *Gramàtica amaziga: Estàndard del diasistema amazic septentrional*. Barcelona: Edicions de la Universitat de Barcelona.
- Naït-Zerrad, Kamal (2011). *Mémento gramatical et orthographique de berbère. Kabyle – Chleuh – Rifain*. Paris: L'Harmattan.
- Pérez-Leroux, Ana Teresa, Alan Munn, Cristina Schmitt, Michelle DeIrish (2004). "Learning definite determiners: Genericity and definiteness in English and Spanish". *BUCLD 28 Proceedings Supplement*. Available at <<https://www.bu.edu/buclid>>. Accessed 05.10.2021.
- Quitout, Michel 1997. *Grammaire berbère: Rifain, tamazight, chleuh, kabyle*. Paris: L'Harmattan. [In: Lamuela (2002)].
- RAE-ASALE (2009). *Nueva Gramática de la lengua española*. Madrid: Espasa.
- Ritter, Elisabeth (1991). "Two functional categories in Noun Phrases: Evidence from Modern Hebrew". In: S. Rothstein (ed.). *Syntax and Semantics 25: Perspectives on Phrase Structure: Heads and Licensing*. San Diego: Academic Press, 37-62.
- Roca, Francesc (1997). *Determinación y modificación nominal en español*. PhD dissertation, Universitat Autònoma de Barcelona.
- Roca, Francesc (2005). *L'ucrainès: Estudi comparatiu entre la gramàtica del català i la de l'ucrainès*. Barcelona: Generalitat de Catalunya, Departament de Benestar i Família.
- Rosado Villegas, Elisa (2007). *La adquisición del sintagma determinante en español como lengua segunda y lengua extranjera* [Tesis doctoral]. Universidad de Barcelona.
- Sarrionandía, Pedro Hilarión (1905). *Gramática de la lengua rifeña*. Tanger: Imprenta Hispano-arábiga de la misión católica. [In: Lamuela (2002)].
- Szabolcsi, Anna (1987). "Functional categories in the noun phrase". In: I. Kenesei (ed.). *Approaches to Hungarian 2*. Szeged: JATE, 167-189.
- Szabolcsi, Anna (1994). "The noun phrase". In: K. É. Kiss, F. Kieffer (eds.). *The Syntactic Structure of Hungarian: Syntax and Semantics 27*. San Diego: Academic Press, 179-275.
- Pérez-Leroux, Ana T., Alan Munn, Cristina Schmitt, Michelle DeIrish (2004). "Learning definite determiners: Genericity and definiteness in English and Spanish". *BUCLD 28 Proceedings Supplement*. Available at <<https://www.bu.edu/buclid>>. Accessed 05.10.2021.

- Svenonius, Peter (2004). "On the edge". In: D. Adger, C. de Cat, G. Tsoulas (eds.). *Peripheries: Syntactic Edges and Their Effects*. Dordrecht: Kluwer Academic, 259-287.
- Valois, Daniel (1991). *The Internal Structure of DP*. PhD dissertation, UCLA.
- Vilosa, Marta (2011). *El español L3 de hablantes de árabe con inglés como L2, un estudio experimental sobre los adjetivos*. Master dissertation, Universidad de Jaén.
- Vilosa, Marta (2021). *La adquisición del SD en español por adultos de lengua materna árabe egipcio*. PhD dissertation, Universitat de Girona.
- Vycichl, Werner (1989). "Article défini". In: *Encyclopédie berbère*. Vol. VI. Aix-en-Provence: Edisud, 939-940.
- White, Lydia (2003). *Second Language Acquisition and Universal Grammar*. Cambridge: Cambridge University Press.
- Zamparelli, Roberto (2000). *Layers in the Determiner Phrase*. New York: Garland.

### **Acknowledgements**

Authors in alphabetical order. This research has been partly supported by the Ministerio de Economía y Competitividad (FFI2017-87140-C4-2-P) and the Generalitat de Catalunya (2017SGR1194). We thank two anonymous reviewers for their comments and suggestions on the previous version, and Suzy Beltagui and Anna Stern for their help with Egyptian Arabic. We also thank all the participants in the study.

Carla Ferrerós Pagès  
ORCID: 0000-0001-6048-3604  
Universitat de Girona  
Departament de Filologia i Comunicació  
Pl. Ferrater i Mora, 1  
17004 Girona  
Spain  
carla.ferreros@udg.edu

Francesc Roca Urgell  
ORCID: 0000-0001-9186-4899  
Universitat de Girona  
Departament de Filologia i Comunicació  
Pl. Ferrater i Mora, 1  
17004 Girona  
Spain  
francesc.roca@udg.edu

Marta Vilosa Sánchez  
ORCID: 0000-0002-2471-7837  
École Supérieure Libre des Sciences  
Commerciales  
Km 22 of Cairo-Alexandria Desert Road  
Giza  
Egypt  
martavilosa@gmail.com



## **Information for Contributors**

*Beyond Philology* is an international journal of linguistics, literary studies and language acquisition. The journal publishes articles, reviews, reports and interviews. The language of the journal is English.

The preferred length of papers is 3500 words, but shorter and longer manuscripts are also considered for publication. Manuscripts should be accompanied by abstracts of about 150 words and keywords.

Each contributor should send a note stating specifically that his/her manuscript is only being considered for publication in *Beyond Philology*.

Manuscripts (papers, book reviews, reports, interviews) and books submitted for review should be sent to:

beyond.philology@ug.edu.pl

For more information see:

[https://fil.ug.edu.pl/wydzial\\_filologiczny/instituty/institut\\_anglistyki\\_i\\_amerykanistyki/czasopismo\\_naukowe\\_beyond\\_philology](https://fil.ug.edu.pl/wydzial_filologiczny/instituty/institut_anglistyki_i_amerykanistyki/czasopismo_naukowe_beyond_philology)

<https://czasopisma.bg.ug.edu.pl/index.php/beyond>

All correspondence concerning subscriptions should be sent to:

Wydawnictwo Uniwersytetu Gdańskiego  
ul Armii Krajowej 119/121  
80-824 Sopot, Poland  
Phone/fax: (48) 58 523 11 37  
<http://www.wyd.univ.gda.pl>  
Email: [wyd@ug.gda.pl](mailto:wyd@ug.gda.pl)

