Differential Object Marking in Corsican

Distribution, triggers, functions.

Inauguraldissertation

Zur Erlangung des Grades eines Doktors der Philologie

dem Fachbereich Philosophie und Geisteswissenschaften der

Freien Universität Berlin

vorgelegt von Kathrin Anne Neuburger

aus Berlin

Berlin 2020

Erstgutachterin: Prof. Dr. Judith Meinschaefer, FU Berlin

Zweitgutachter: Prof. Dr. Antonio Fábregas, UiT Tromsø

Tag der Disputation: 21.12.2020

Acknowledgements

Content

List of abbreviations

1sg, 1pl	first person singular, first person plural
2sg, 2pl	second person singular, second person plural
3sg, 3pl	third person singular, third person plural
a-marking	DOM-marking
ADECEC	*L'Association pour la Défense de l'Étude de la Langue Corse de l'Est et du Centre*
AGR/AgrP	Agreement/Agreement Phrase
ALEIC	*Atlante Linguistico Etnografico Italiano della Corsica*
ALF C	*Atlas Linguistique de la France: Corse*
ARC	*Action Régionaliste Corse*
Asp	Aspect
BDLC	*Banque de Données Linguistique Langue Corse*
CNED	*Centre National d'Enseignement à Distance*
CNRS	*Centre National de la Recherche Scientifique*
coin	coincidence
CompV	Complementiser position of VP
CP	Complementiser Phrase
CTC	*Collectivité Territoriale de Corse*
C_{Lg}	Language specific categories
C_{UG}	Categories of the Universal Grammar, universal categories
D/DP	Determiner/Determiner Phrase
D°	Determiner Head
DatP	Dative Phrase
DET	determiner
DO	Direct Object
DOM	Differential Object Marking
ECP	Empty Category Principle
F	Functional (element)
FP	Functional Phrase
FLNC	*Fronte di Liberazione Naziunale Corsu*
ident	identity (feature)
ind	referential index
IndP	Indexical Phrase
INFL	Inflection
INSEE	Institut national de la statistique et des études économiques

InflP	Inflection Phrase
IO	Indirect Object
K	Case
KIP	Kind Determiner Phrase
KP	Case Phrase
L1	first language/mother tongue
L2	second language
N	Noun
NALC	*Nouvel Atlas Linguistique et Ethnographique de la Corse*
NE	Nominal Element
NP	Nominal Phrase
num	number
OT	Optimality Theory
OV(S)	Object-Verb-(Subject)
PDP	Predicative Determiner Phrase
PL	Plural
PP	Preposition Phrase
Pred / Pred	Predication (feature)
QC	Quantifier Construal
QR	Quantifier Raising
Refl	reflexive
SpecDP	Specifier of Determiner Phrase
SDP	Strong Determiner Phrase
SG	singular
T	Tense
TP / TENSEP	Tense Phrase
UG	Universal Grammar
V	Verb
*v*P, VP	Verb Phrase
WWI, WWII	World War One, World War Two
ι	referential index
φ/φP	phi/phi Phrase
Φ-features	Semantic features of person, number, gender, and case
#P	Number Phrase

1 Introduction

Differential Object Marking (DOM), in its most basic and traditional conception, is the marking of certain, but not all, nominals in direct object (DO) function (in nominative–accusative languages, the accusative object) according to different semantic, syntactic and pragmatic criteria by means of an overtly expressed morphosyntactic marker. In Romance, many varieties show DOM; for example: Spanish (cf. RAE 2009: 2638, § 34.9a; von Heusinger & Kaiser 2003, 2005, 2007); Portuguese; Galician; many Italo-Romance varieties, such as Sicilian (Iemmolo 2011), Sardinian (Jones 1995), or Neapoletan (Ledgeway 2011); Occitan varieties; Romanian (von Heusinger & Chiriacescu 2011); some Rhaeto Romance varieties; and Corsican. While Romance DOM systems may show distributional overlaps – for example, in many varieties personal pronouns or [human] proper nouns are overtly marked in DO function – other systems pattern in a very specific way or display at least specific characteristics. Hence, even if many Romance varieties have DOM, no Romance DOM system, to the best of my knowledge, is identical to any other.

In Corsican, DOM is expressed by the morpheme *à*, which is homophonous to the dative and locative prepositions *à*. It appears with the DO if certain semantic (i.e. humanness) and/or syntactic conditions apply, and does not surface if these do not apply. The aim of this thesis is to give a deeper insight into the patterns and functions of DOM in this Italo-Romance variety, based on empirical data gathered during two fieldtrips during the summers of 2010 and 2011.

Corsican DOM patterns in a rather complex way. In the literature we find two major generalisations: on the one hand, Corsican DO-nominals get DOM-marked, i.e. *a*-marked based on animacy or 'personification'; on the other hand, the presence of a definite article prevents the *a*-marking of common nouns and (certain) proper nouns. Hence, both semantic and syntactic properties are relevant and reveal a complex marking system.

This thesis aims to give a detailed description and analysis of how these two properties – animacy and definiteness – may interact (or not) on a semantic and syntactic level. Traditionally, such complex mixed DOM systems, whereby DOM is driven by and related to different properties, have been considered rare, or even impossible, tue to an absence to the contrary. Usually, either semantics (predominantly, but not exclusively, animacy) or syntax (definiteness and specificity, rather than animacy) are the trigger for DOM. To analyse Corsican field datat collected for this study, I used well-known DOM theories, such as disambiguation theories and the transitivity hypothesis, as well as syntactic accounts from the generative framework.

Alongside the theoretical embedding of Corsican DOM, my field data forms an essential part

of this thesis. To date, Corsican lacks independent and reliable language documentation. Of the many reasons for this, I will mention here only some: generally, Corsican has a predominantly oral tradition, and any written documentation is scarce – outside Corsica it is difficult to find any books, newspapers, texts, etc. written in Corsican; also, finding informants is a difficult endeavour. Due in particular to severe language political regulations in the first half of the 20th century, it was and it is still difficult for any 'outsiders' to establish contact with speakers of Corsican. This is even more the case with finding unbiased informants, i.e. persons who are not 'language activists' of some kind but simply speak Corsican in their everyday lives. Hence, collections of spontaneously spoken Corsican neither exist, nor are they easy to compile.

Second, compared to other Romance varieties, Corsican is rather underrepresented in linguistic studies. This may be due to its 'jammed' position – belonging linguistically to the Italo-Romance sphere, but politically to France. Corsican has been largely neglected by linguists working on Italian varieties or French – what attention it has received has been mainly from Corcian linguistcs and a few other exceptions. Moreover, Corsican language data and its descriptions are rarely embedded into a theoretical framework.

Third, and as a consequence of the aforementioned issues – already being a scarcely documented variety – Corsican (morpho)syntax is a rather neglected field of linguistic studies (compared to phonetics, phonology or lexis). This is indicated, not only by the small number of explicitly morphosyntactic studies, but also by the often-made claim that Corsican syntax is diatopically homogeneous, as well as the lack of an adequate tool, such as a corresponding linguistic reference corpus that allows for morphosyntactic analyses.

In contrast to this scarcity of sources and studies, there is a longstanding tradition of descriptive accounts on Corsican DOM. Compared to other morphosyntactic phenomena, it is the most frequently mentioned phenomenon. Nevertheless, many of these studies remain somewhat sketchy, restricting themselves to the enumeration of just a few examples, accompanied by isolated remarks on the relevant nominal categories. Hence, the picture of Corsican DOM in the literature is fragmentary, as the nominal classes mentioned are not backed up by (sufficient) data to validate the descriptions, nor do the data always match the mentioned classes. This existing body of data provides only limited (if any) information about origin and methodology. Hence, the studies do not follow current 'good practice' for empirical studies, in terms of replicability and reliability.

Based on the available information, I identified two main goals for this study: on the one hand, to enhance the value of Corsican as a topic of scientific studies by providing new empirical data that

will allow for new approaches – in particular, studies on morphosyntax; on the other hand, by analysing the data against the theory on (Romance) DOM systems, i.e. different semantic and syntactic approaches, to classify Corsican DOM among the Romance DOM-systems in general. To achieve these goals, I conducted two fieldtrips to collect empirical language data, and analysed these data against some of the most common theoretical approaches to Romance DOM in semantics and syntax. I am interested in how the Determiner Phrase (DP) and the nominal (i.e. structure and item) interact, and how they influence each other regarding the phenomenon of DOM. In addition, I provide background on Corsican language and culture, including political implications and regulations, for a better understanding of the data (new an pre-existing) with regard to quality and quantity. I will also describe the state of the art regarding existing linguistic studies about Corsican DOM.

The dissertation is structured as follows:

In Chapter 2, I present DOM in Romance from different perspectives. First, I introduce the main conditioning features, i.e. animacy, definiteness, individuation, specificity and (cor)related concepts. The differences between 'inherent' and 'discourse-related' features and their possible impacts are also discused in this section. Definitions (section 2.1.1) and formal representations (section 2.1.2) are also offered.

I consider traditional semantic accounts on (Romance) DOM in the next section (2.2), in the light of the available information about features – starting with some remarks on markedness, prototypicality and iconicity (section 2.2.1), followed by different accounts on disambiguation (section 2.2.2) and transitivity (section 2.2.3).

In the third theoretical section (2.3), I present formal syntactic accounts regarding the DP-structure and DOM. I start with a section about the DP-structure (section 2.3.1), considering different split approaches to the DP. These do not actually relate to DOM in particular, but offer a detailed insight into the syntactic structure of the nominal domain, in particular the so-called 'left edge' of the DP, regarding specifiers and D-heads (D°s). After this theoretical grounding, I continue with specific accounts on DOM, most of them based on Spanish data as this is the most-frequently studied Romance DOM system from the perspective of formal syntax. Section 2.3.2 concerns so-called 'Dative accounts'. Their authors relate DOM-marking, and in particular the *a*-marked accusative object, to the structures of dative marking. The other section on syntactic accounts on DOM concerns so-called 'Accusative accounts' (section 2.3.3). The authors of these regard both *a*-marked and unmarked DOs as accusative objects. They explain overt *a*-marking as a result of specific Case

checking/assignment processes that only concern specific DOs.

Chapter 3 is dedicated to the Corsican language in a broad sense. The first part is dedicated to language differentiation and classification (section 3.1), and addresses diatopic differentiation (section 3.1.1), language classification (section 3.1.2) and the question of what is considered 'original' Corsican (section 3.1.3). In terms of the history of science and Corsican linguistics as an (independent) discipline, one more section is dedicated to sources used to study Corsican (section 3.1.4), and another to an overview on linguistic studies dedicated to Corsican DOM (section 3.1.5). The second part of the third chapter is dedicated to the history of Corsican settlement and language, and allows for a better understanding of Corsican and its genesis, but also its social-historical and cultural embedding as an oral language variety from pre-Tuscan times until the 21st century (sections 3.2.1–3.2.4).

Having thus introduced Corsican in a more social-cultural sense, the focus of Chapter 4 is on the state of the art regarding Corsican DOM. Based on previous studies, I look at different nominal classes and which ones display DOM (section 4.1), and at the features the authors have considered, in particular humanness (section 4.2). Inspired by studies on Romance DOM, I also consider verb semantics and potential interrelations with DOM (section 4.4). Corsican DOM has rarely been embedded into any theoretical framework to explain its pattern, the exceptions being where disambiguation accounts have been considered; these are presented in section 4.5. Some remarks on the most current theory on Corsican DOM, Giancarli's *Hypothèse de l'entonnoir*, are presented in section 4.6. Sections 4.1 to 4.6 are concerned with accounts based on the description of DOM. By contrast, section 4.7 is dedicated to prescriptive rules regarding Corsican DOM, as found, for example, in textbooks or school grammars. Section 4.8 gives a summary of all the findings presented in Chapter 4. From the information presented I distill my research questions, which are answered in the final section (7.3).

In Chapter 5, I talk in detail about my methodology for gathering empirical language data, and present my fieldwork undertaken in Corsica. I start with the theoretical background, i.e. the differences between grammaticality and acceptability (section 5.1), and go on to present the chosen methodology, i.e. forced choice and evaluation (section 5.2). In section 5.3, I present my questionnaire, how it was designed, and why. I conclude Chapter 5 with a description of my fieldtrips. This last part is divided into three sections: I write about the experiences of other linguists doing fieldwork in Corsica (section 5.4.1), my own fieldwork (section 5.4.2), and finally I conclude the Chapter with an overview of my data and informants (section 5.4.3).

Chapter 6 is dedicated to the results from the questionnaire. I present the results according to

the nominal classes the items belong to: starting with kinship terms (section 6.1); followed by proper nouns with different animacy values (section 6.2); common nouns (section 6.3); and finally proforms, i.e. different kinds of pronominally used items, (section 6.4). The chapter is concluded with a summary (section 6.5).

In Chapter 7 I discuss the results of Chapter 6 against the theoretical implications set out in Chapter 2. I show that the well-known features of animacy (or rather humanness) and definiteness (or probably referentiality) have also an impact on Corsican DOM-marking (section 7.1). In relation to this I look at certain nominals in particular, such as e.g. indefinite and negative proforms and common nouns, as they seem to 'fall out' of the DOM-pattern in terms of features. For the second part of the discussion, the incompatibility of the definite article and the *a*-marker takes centre stage (section 7.2). I analyse this incompatibility from different points of view: from the perspective of semantic and syntactic accounts (sections 7.2.1 and 7.2.2, respectively); and also from the perspective of splitting up the Corsican DP, and in particular the DOM-DP (section 7.2.3). As indefinite and negative quantifiers pose a problem to many accounts on DOM in terms of semantics and syntax, I dedicate section 7.2.4 to these items in Corsican. Finally, I broaden the view from the DO-nominals and the DP-domain, to accusative case assignment and DOM-marking in more general terms (section 7.2.5). I conclude the discussion with a summary, and with answers to the research questions set out in section 4.8.

The thesis is concluded in Chapter 8. The appendix, (Chapter 9), contains the questionnaire, details about the sentences, as well as more detailed information about the informants included in the survey.

Before I start, I want to make a remark on the Corsican language data: the spelling in the examples corresponds always to the original writing in the respective sources. Due to the practices of polyphony and phonemic orthography, Corsican spelling can be very heterogeneous.

2 Factors conditioning DOM in Romance (and beyond)

DOM is a morphosyntactic phenomenon that is related to the characteristics of the nominal in DO function and its context: lexical, semantic referential, morphosyntactic, discourse-pragmatic features, and features related to the information structure, may trigger the realisation of DOM. Animacy and definiteness of the DO referent are considered most relevant (cf. Bossong 1985, 1991, 1998; Aissen 2003; Heusinger & Kaiser 2003, 2005, 2007; Næss 2004, 2007; Malchukov 2005, 2006, 2008; de Hoop & Malchukov 2006; de Swart 2006, 2007; García García 2007; de Hoop 2009; among many others).

Studies on Romance DOM are almost as manifold as imaginable: countless in number, they regard the phenomenon from diachrony and synchrony, within one language, or under a cross-linguistic or typological perspective. The phenomenon has been analysed within different theories, such as Generative Grammar, Construction Grammar, Optimality Theory or Functional Typology. Some studies have considered DOM in relation to other (correlated) phenomena, such as clitic doubling. Nevertheless, most studies concentrate on semantic interpretative aspects, such as animacy and referentiality (see section 2.2).

DOM investigations from different linguistic branches have contributed different results: accounts that consider phonological or prosodic aspects are quite rare: ermarks on DOM related to prosody have been named by Mardale (2008); Iemmolo (2011) mentions the relationship between prosodic stress and topicality with relation to DOM; and Wandruszka (2011) studies Spanish DOM-marked personal pronouns from a prosodic point of view. Semantic accounts probably have the most longstanding tradition among all DOM studies. These are presented in section 2.2. In the last three decades morphological aspects have been considered more often, particularly in relation to (formal) syntactic accounts. In section 2.3 I go into detail about the syntactic accounts, concentrating particularly on the DP-structure and its immediate environment.

Before presenting different accounts, I introduce the two most relevant features, animacy and definiteness, including remarks on their conceptual organisation, such as *values* and *hierarchies* (in section 2.1). Related concepts, such as referentiality, individuation and personification are included in the description. In this way, section 2.1 offers the conceptual background for the subsequent sections on semantic (section 2.2), but also syntactic (section 2.3), accounts on DOM.

I refer to animacy and definiteness as features that have a range of values, i.e. [human], [animate], and [inanimate], or [definite] and [indefinite], respectively. These features and values

contribute to the broader concepts of individuation, personification, and referentiality, (among others). At the same time, animacy and definiteness may also be conceived as broader concepts to which other features contribute, such as uniqueness or familiarity contributing to definiteness, or person contributing to animacy.

Conceptually, the values are conceived as simple non-binary values in privative opposition, e.g. a definite nominal is considered [definite], while an item not bearing the value is unmarked, i.e. not specified for definiteness. For example, in English, the common noun *man* is not specified for definiteness. If the noun appears with the definite article, *the man*, it becomes definite, while *a man* will be interpreted as indefinite. The non-specification of a feature allows for different interpretations. It may capture *optionality.*[1]

In a study on markedness of objects, Næss (2004) explores the idea of non-specification of features. She departs from an "inclusive asymmetry" (Næss 2004: 1196), based on a privative opposition, (see Table 1). This kind of opposition operates with a single feature, or one value of a feature, and its presence and absence are opposed: 'A versus ?'. For '?', there are three possible interpretations: (i) the absence of A, (ii) the non-specification for A, or (iii) the contrary of A.

Table 1: Kinds of oppositions following Næss (2004)

Kinds of oppositions	Pattern	Interpretation of values	
polar opposition	A versus B	value 1 specified for A value 1 = A	value 2 specified for B value 2 = B
privative opposition ("inclusive asymmetry", Næss 2004: 1196)	A versus ?	value 1 specified for A value 1 = A	value 2 not specified value 2 = ? (i) absence of A (ii) non-specified for A (iii) contrary of A

The concept of an 'optional range', in terms of the opposition-patterns given in Table (1), is best captured by a non-specification of the second value; it leaves open all three interpretations (cf. Næss 2004: 1197). In terms of DOM, that means "[…] what is encoded by DOM is a markedness relation of inclusive asymmetry – the specification versus nonspecification of a property […]" (Næss 2004: 1197). Hence, the marked DO is specified for a feature and/or a certain value, for which the unmarked DO is not specified.

Nevertheless, unmarked items are not always variable in interpretation or function. Rather, items unmarked for a certain feature sometimes bear a specific interpretation and/or a specific

---

[1] Optionality may also be possible if the invoked feature(s) do not capture the marking pattern at all., in which caseother features may be relevant, besides those already mentioned. For more on optionality see section 2.2.1.

function. This can be accounted for in terms of zero marking. Hence, the differences between zero marked and (truly) unmarked items or values influence their interpretation: zero marking corresponds to one interpretation or one value, while truly unmarked items have a variable interpretation.

2.1 Features: definitions and formal representation

First, I present the relevant features, definiteness and animacy, providing their definitions and how they are related to the phenomenon of DOM.

2.1.1 *DOM-Features: definitions*

We can distinguish between properties of the noun that are lexically inherent, and the features given by morphosyntax, i.e. structurally determined values. Properties of the first type are traditionally called 'inherent' (e.g. Bossong 1998); they are independent of the respective linguistic context or function. This holds true for animacy in a common sense. Properties of the second type are often called 'discourse-related' properties, such as definiteness or specificity.

Definiteness is related to the nominal classes a given noun is associated with, and/or the syntactic context in which a given nominal appears. Definiteness occurs in different froms. On the one hand, certain nominal classes are associated with certain definiteness values; for example, proper names or personal pronouns are interpreted as [definite]. On the other hand, definiteness is given by the morphosyntactic context, by the contribution of other items such as definite articles, determiners, possessives, or deictics, and the position of the noun in the syntactic structure. Since Abney (1987), the majority of generative syntactitians agree that definite nouns are located in a DP-structure above the nominal phrase (NP), with definiteness hosted in the determiner head (D°) (among many others, cf. Abney 1987; Longobardi 1994; Giusti 1997; Alexiadou et al., 2007). The syntactic structure of a (split) DP may capture the different manifestations of definiteness.

Bossong (1998), von Heusinger and Kaiser (2003), and de Hoop (2009) all make a distinction between animacy on the one hand and definiteness on the other hand, as inherent and discourse-related features, respectively. All authors account for the difference in similar terms: von Heusinger and Kaiser (2003) consider animacy as a "lexical or cognitive feature of NPs", while "definiteness is a discourse-pragmatic property of the discourse item representing the DO" (von Heusinger & Kaiser 2003: 63). De Hoop (2009) distinguishes between the two features with the following words: "[W]hile definiteness can be conceived of as a feature related to the syntax of noun phrases, animacy is a feature related to the individuals (in the world or discourse) that the noun phrases refer to" (de Hoop 2009: 235). In summary, animacy is basically an ontological rather than a linguistic feature, related to the nature of the referents of nominals, while definiteness is considered a linguistic feature,

related to the (morpho)syntax of nominals, with structural and semantic interpretative effects.

DOM-systems tend to be either animacy driven or definiteness driven, even if the features interact in many cases. The features display differences in several respects: definiteness driven DOM systems may be sensitive to the syntactic structure of the (un)marked nominal, e.g. Hebrew (cf. Danon 2006), while animacy driven DOM systems are considered to be not, e.g. Spanish. The value [definite] may require or result in a specific syntactic structure of the nominal or nominal phrase (see syntactic accounts on DOM in section 2.3), while animacy-based DOM systems do not: "a syntactic difference between NPs and DPs could explain a difference in case marking in relation to definiteness, but not for animacy. Animacy is usually conceived of as a semantic feature rather than a syntactic one" (de Hoop 2009: 241). Contrary to de Hoop (2009), Jones (1999) proposes an account that unifies the structurally conditioned and animacy-driven Sardinian DOM-pattern under a syntactic approach. A similar pattern is also found in Corsican: Corsican DOM seems to be basically animacy driven, and at the same time sensitive to the syntactic structure of the DO in question (see Chapter 4).

With regard to the meaning and the interpretation of nominals, marking triggered by definiteness may result in a different definiteness value, and interpretation of the marked nominal as 'definite', 'specific' or the like. Marking due to animacy does not change the actual animacy value of the nominal's referent (see also de Swart 2006, 2007, for this differentiation).

For this thesis, I proceed from both a semantic interpretative and a syntactic concept of definiteness: definite nouns are semantically [definite] and they are syntactically DPs. Definiteness may be conceived with common sense or world knowledge (e.g. the use of proper nouns) or it may be related to the linguistic discourse (e.g. referents of definite common nouns that have been previously introduced to the discourse). New referents are introduced to the discourse by indefinite nouns. With regard to animacy, I proceed from an ontological feature with the respective values [human], [animate] and [inanimate], without further syntactic implications.

2.1.2 *DOM-features: formal representation*
So far, I have presented the relevant features, their definitions and their relationship to DOM. I have mentioned their values but said nothing as to how these values are organised – in relation to each other, to the superordinated structures (e.g. noun classes), or the syntactic structure of nominals and nominal phrases. To adress this gap, I refer to hierarchies. In section 2.3.1, I present split DP-structures and the resulting impacts of such approaches.

In many morphological and syntactic phenomena, animacy and definiteness have a role to play. Helmbrecht and Handschuh (2016) list, among the animacy driven grammatical phenomena:

categorical differences in third person pronouns (human vs. non-human), distinctions in interrogatives (human vs. non-human), DOM, plural marking on nominals, split ergative/accusative case marking patterns, verb agreement, passive and antipassive constructions, and so-calles 'fourth person'.

Many studies have elaborated and applied hierarchies of animacy and definiteness. In these hierarchies, nominals are ordered according to: animacy values and definiteness values, the nominal classes they belong to, and distinctions based on person. The values of animacy, definiteness and person are hierarchically ranked (see Table 2):

Table 2: Features of the nominal hierarchies and the ranking of their values.

feature	*ranking of values*
person	1st/2nd < 3rd
noun class	proforms < proper nouns < common nouns
animacy	human < animate < inanimate

The relative position of a single nominal element (NE) allows for predictions about other NEs included in the hierarchy. Most commonly the scales are conceived as implicational scales. If the phenomenon in question, here DOM, affects a certain nominal type within the hierarchy, items of other types with a 'higher' value of the feature, will be affected as well; the further left within the scalar representation a nominal is classified, the more likely it is to get DOM-marked, and the other way around (cf. Aissen 2003: 436). In other words, if [animate] nominals are marked by DOM, [human] nominals also get DOM-marked; if proper nouns are marked, personal pronouns also bear the marking.

The ranked nominal classes are represented in hierarchies. These hierarchies can have different labels, such as referentiality (e.g. Silverstein 1976),[2] definiteness (e.g. Aissen 2003), or animacy (e.g. Croft 2003). Regardless of the label, these hierarchies have a property in common: to define a certain position within the hierarchy, they combine values of the aforementioned features – animacy, definiteness, person, specificity, and the noun class. As an example, see Croft's (2003) hierarchy, (Table 3).

---

[2]   Silverstein (1976) proposed the hierarchy to show the pattern of person splits with regard to accusative–ergative systems.

Table 3: Extended[3] animacy hierarchy (Croft 2003: 120; following Dixon 1979: 85 and Silverstein 1976)

1st/2nd person pronoun < 3rd person pronoun < proper name < human common noun < non-human animate common noun < inanimate common noun

From the position of a nominal within the hierarchy, Croft predicts how likely a noun is to bear a certain syntactic function: nominals on the left edge of the hierarchy are more likely to function as subjects,[4] while nominals at the right edge are more likely to be DOs.

With regard to DOM, its many patterns are not captured in an unequivocal way by the existing hierarchies. The hierarchy may be too 'rough': relevant features may not be included in the hierarchy, or "different combinations of values [may] have the same status in the hierarchy", but pattern in a different way (Fábregas 2013: 10). Laca's (2006) hierarchy tries to capture the equivalence of values and value combinations (see Table 4).

Table 4: 'Explicit, integrated hierarchy' following Laca (2006)

I >	II >	III >	IV >	V >	VI >	VII
human pronoun	human proper name, animate pronoun	human definite, animate proper name, inanimate pronoun	specific human indefinite, definite animate, inanimate proper name	human non-specific, indefinite specific animate, definite inanimate	animate non-specific, inanimate indefinite specific	inanimate non-specific

Human and animate entities usually bear one value of animacy based on ontological considerations, whereas the classification of inanimate entities may be rather fuzzy. This is where Yamamoto's (1999) idea of 'inferred' animacy comes in: he considers entities with an inferred animacy value, i.e. the interpreted value diverging from the ontological value, to be 'borderline' cases. Among these he quotes, for example, machines (cars, computers) addressed by 2nd person pronouns; but also organisations, geographical entities, and local communities, bearing proper names. While addressing

---

3   "Since semantic animacy is only part of the hierarchy, we will call it the extended animacy hierarchy" (Croft 2003: 130).

4   This statement is supported by frequency effects: cross-linguistic studies show that the subject function is more frequently fulfilled by nominals with a human referent, as opposed to DO referents that are mostly inanimate (for an overview of the results of cross-linguistic studies, see Iemmolo 2011: 28 ff.).

machines by 2nd person pronouns representsinstead an exceptional use of these pronouns, addressing locations or organisations by proper names is cross-linguistically the dominant pattern. Hence, Yamamoto's 'borderline' cases diverge strongly with regard to quantity. Inferred animacy is linguistically perceivable when, for example, human and/or animate characteristics are attributed to inherently inanimate objects, such as attributing agency to meteorological phenomena, or to events that are referred to by proper nouns (cf. Yamamoto 1999: 23).

The other main trigger, definiteness, displays two values, [definite] and [indefinite]. Within the hierarchies used in our context, definiteness is related to *specificity*. As an example, let us consider the so-called 'definiteness' or 'referentiality scales' given by Aissen (2003), von Heusinger and Kaiser (2003 and *passim*), and many others thereafter (see Table 5).

Table 5: Referentiality scale (cf. Aissen 2003: 437; modification by von Heusinger & Kaiser 2005: 38 and me)

personal pronouns > proper nouns > definite NPs > indefinite specific NPs > non-specific NPs				
personal pronoun	proper noun	NP [definite]	NP [specific]	NP [non-specific]
definite			indefinite	
specific				unspecific

Von Heusinger and Kaiser (2003) conceive both features, definiteness and specificity, as discourse pragmatic or semantic referential notions that contribute to the referentiality of a nominal. Definiteness is defined by von Heusinger and Kaiser (2003) in the following terms:

> [D]efiniteness is a discourse-pragmatic notion that indicates that the discourse referent associated with a definite expression can be identified with an already introduced discourse item. Thus, definiteness does not express the identifiability of the referent (in the world) […] Rather, definiteness expresses familiarity in a discourse structure. The discourse structure is understood as an intermediate structure between the linguistic expression and their referents in 'the world'. (von Heusinger & Kaiser 2003: 42, 43)

Von Heusinger and Kaiser (2003) argue that definiteness and specificity are independent, as definite nominals can also be used in a non-specific way. This is illustrated by the following example (1), where *the book* is [definite] but at the same time [non-specific]:

(1) *Busco   el   libro   en   el   que   se   analice   mejor   el   modo*

SEARCH   the   book   in   the   that   REFL   analyse   better   the   modus

en   las   oraciones   de   relativo.

in   the   clauses   of   relative

'I'm looking for a book in which the modus of relative clauses is better analysed.'

(cf. von Heusinger & Kaiser 2003: 50 (23))

For von Heusinger and Kaiser (2003), the interpretation of a nominal as [specific] is not given by the nominal itself, but by an element given in the (broader) linguistic context to which the nominal in question is related: "[T]he specific expression is linked or anchored to another expression (the anchor) and therefore its interpretation is independent of the direct linguistic context. The interpretation depends instead on the interpretation of the anchor expression" (von Heusinger & Kaiser 2003: 48). For example, *them* in the following sentence (2) is linked or anchored to the phrase *Anna and her friends*:

(2) *Yesterday I met [Anna and her friends]$_i$ in the theatre. I invited [them]$_i$ for my birthday party next week.*

In the context of Spanish DOM, we can state that not all *a*-marked nominals are specific. According to López (2012: 19), *a*-marked DOs may introduce new discourse referents; they need not necessarily be specific. But specificity requires the referent of a specific nominal to be introduced at some other point in the discourse.

Jaeggli (1982) offers a more traditional and basic semantic morphological idea about the notion of specificity. Very roughly, he states that *a*-marking makes a difference in interpretation with regard to specificity in the context of indefinites: *a*-marked indefinite DOs are interpreted as specific, while unmarked indefinites have no specific interpretation. For him, the DOM-marker is a specificity marker. Diesing (1992), among others, provides subsequent approaches to specificity, elaborating on this basic notion. She assumes the specificity interpretation to be related to a specific structural configuration which includes DOM. Hence, while particular-specific references may result in *a*-marking and, vice versa, generic references never result in *a*-marking.

The semantic pragmatic value of specificity is emphasised by Stark (2006). Her idea is based

13

on two different kinds of reference: a nominal can either refer generically, i.e. to a type, or particularly, to a subset or to a singleton set. Within the latter (i.e. subsets and singleton sets), Stark distinguishes between specific und non-specific reference. The semantic referential notion comes along with certain structural differences. As with the other authors mentioned above, for Stark too, the specific interpretation is related to the discourse-level, to an invariable reference:

> Die Unterscheidung in *spezifische* und *nicht-spezifische* Interpretation von Nominalsyntagmen betrifft im wesentlichen eine Unterscheidung in 'ein bestimmtes X' und 'irgendein X', also die Unterscheidung in bestimmte und nicht-variierende Referenz in der Textwelt […] und beliebige oder variierende Referenz. […] Es ist […] der Kontext, vor allem die Bedeutung des Prädikats für seine Argumente, aber auch die interne Struktur des Nominalsyntagmas oder die Anwesenheit bestimmter Operatoren […], die den Unterschied in der Interpretation der fraglichen Nominalsyntagmen bedingen. (Stark 2006: 42, 43, emphasis theirs)

> [The distinction between *specific* and *non-specific* interpretation of nominal phrases is basically related to the distinction between 'a specific X' and 'any X', hence, the distinction between certain and non-varying reference in the discourse […] and arbitrary and non-varying reference. […] It is […] the context, in particular the meaning of the predicate to its arguments, but also the internal structure of the nominal phase or the presence of certain operators […], which determine the different interpretations of the nominal phrase in question; KAN]

Another aspect of definiteness, *quantification*, is illustrated by Hopper and Thompson (1980) in terms of specificity. To illustrate the nexus between definiteness and referentiality they give the following examples (3):

(3)   a. Fritz drank the beer.

        b. Fritz drank some beer.   (Hopper & Thompson 1980: 253)

In the first example (3a), Fritz has drunk all the beer available to him in a given situation: the determiner expresses a specific reference and all-quantification. Contrary to that, in the second example (3b), the indefinite quantifier *some* does not express specific reference. It refers to a unspecific subset, not further defined. By implication, we know that there is more beer available in

the given situation than the quantity drunk by Fritz. A lower degree of individuation correlates with a lower degree of referentiality. Hopper and Thompson (1980: 288) consider referentiality the most relevant among all mentioned features contributing to individuation (see also section 2.2.3, and in particular, Table 7).

According to Lyons (1999), definiteness in its semantic and pragmatic core is universal. It refers to 'identifiability' in the broad sense. We find an equivalent in Wiltschko's (2014) identity feature, which contributes to nominality (Wiltschko 2014: 188 ff.; see also section 2.3.1). But the morphosyntactic and grammatical manifestations of definiteness are language-specific. Definiteness may either appeal to a *linguistic* context – the discourse – and has anaphoric function, or it appeals to a *non-linguistic* context, where it includes world knowledge or knowledge of the situation related to the utterance (cf. Lyons 1999: 158 ff.). For example, in English we associate anaphoricity, familiarity, general knowledge and reference to specific situations, with definite DPs.

In the section above, I have introduced the main semantic features, animacy and definiteness, elaborating their linguistic and ontological values. Major focus was given to their hierarchical organisation and their interplay with other features, e.g. specificity.

2.2 Semantic approaches on DOM

Semantic accounts of DOM are driven by two lines of thinking: differentiation of nominals in different functions, such as subject and DO function; and/or highlighting or indexing of these nominals. The idea of interpreting morphological marking as a distinguishing means has a long tradition in linguistics (e.g. since Diez 1836; Meyer-Lübke 1900). DOM offers a *morphological differentiation* of the arguments and their syntactic roles, different from other distinguishing strategies, such as word order in French. In this sense, DOM is employed "when both agent and direct objects share the same semantic properties, and other clues for identifying the NP's syntactic role are not found in the clause" (Iemmolo 2011: 26).

By contrast, the idea of interpreting overt marking as a means to *highlight* the most typical element has systematically broken ground since Hopper and Thompson's (1980) transitivity hypothesis. The *indexing* function of the marker is emphasised in contexts where the verb is in the centre of investigation, i.e. the indexing radiates from the verb (cf. Malchukov 2005; Schikowski & Iemmolo 2015). Nevertheless, this distinction is not consistently reproduced in the literature. Indexing may also refer to the function of the morphological marker to *indicate* or to *highlight* the marked item due to the nominal's properties. In this sense I will use the termin in this thesis.

Before going into semantic approaches, I will first discuss relevant concepts, such as

markedness, prototypicality and iconicity.

2.2.1 *Markedness, prototypicality and iconicity*

Markedness is a vast and complex issue, related to various linguistic phenomena (Næss 2007: 5) and used in different linguistic frameworks. It originates in the context of Prague structural linguistics, where a distinction is made between 'marked' and 'unmarked', which corresponds to the privative opposition between 'bearing a feature' or 'not bearing a feature' (see also introductory section of Chapter 2, Table 1 on oppositions).

Mayerthaler (1981) relates markedness to naturalness: an unmarked structure is more natural than a marked structure. A structure is considered natural with respect to frequency, language acquisition, and language change: if a structure is more frequent within a language or a language family, and/or the structure is acquired relatively early, and/or it is either quite resistant to language change or often the result of such a change, it is considered unmarked (cf. Mayerthaler 1981: 2). Additionally, Mayerthaler (1981) assumes an iconic relationship between semantic and morphological markedness, i.e. semantic unmarked items are expressed by morphologically simpler, or less complex structures.

Comrie (1989) also relates markedness to naturalness and employs it on transitivity:

> The most natural kind of transitive construction is one where the A[gent] is high in animacy and definiteness, and the P[atient] is lower in animacy and definiteness; and any deviation from this pattern leads to a more marked construction […] The construction which is more marked in terms of information flow should also be more marked formally. (Comrie 1989: 128)

In this sense, proponents of the disambiguation accounts, such as Bossong (1985, 1991, 1998) or Aissen (2003), follow a traditional markedness account: what is less natural, less normal, is marked, semantically and morphologically. With respect to Romance DOM this may concern DO nominals with, for example, [definite] and [human] feature values. By contrast, Hopper and Thompson (1980), and also Næss (2004), turn this argumentation upside down: what is most typical becomes marked or indexed.

All authors mentioned above agree that a high degree of animacy and/or referentiality triggers overt marking. But from a disambiguation point of view, such as for Bossong (1985, 1991, 1998) and Aissen (2003), a high degree of animacy and definiteness of the object reflects semantic markedness compared to the subject. The morphological markedness corresponds to the semantic

markedness. Markedness expresses the *deviation* from the natural, the normal, the prototypical. For Hopper and Thompson (1980) the *highlighting* function of the morphological marker is to the fore – to underline a high degree of individuation, which is in turn composed of features such as animacy and referentiality, among others (see Table 7, below). As a consequence, the interpretation and evaluation of the marking, or rather the marked DO, in terms of markedness and/or prototypicality, diverge significantly.

For Næss (2004), morphological markedness corresponds to morphological complexity: the more morphemes a construction has, the more complex it is. For example, comparing a DOM-marked DO to an unmarked DO, the former is more complex and therefore considered marked. As such, morphemes or morphological structures in general do not correspond in a one-to-one fashion to semantic features, or more specifically to marked semantic values. The two criteria of semantic and morphological markedness are not automatically mutually dependent.

Apart from markedness, prototypicality is used as a means of evaluation. Following Rosch's (1978) idea on categorisation, linguistic items or structures are 'prototypical' if they are considered a good or even the best representative of a certain category. The category itself is defined by a set of properties, which are weighted differently, i.e. certain properties are considered more important, others less so, and still others not at all important for a given category. Croft (1993) notes that the notion of markedness with reference to prototypicality does not hold true in absolute terms: what might be a marked combination with regard to one prototype, might be unmarked with respect to another. For prototype accounts, belonging to a category is a matter of degree. DOs are prototypical if they display a high degree of objecthood, defined by certain properties (for more detail, see section 2.2.3, Tables 6 and 7).

The typological view brings together markedness and prototypicality accounts: "Typological markedness patterns might be also viewed as instantiations of prototypical categories" (Iemmolo 2011: 8). In particular, Croft (1993) applies the markedness pattern to evaluate whether a linguistic item or structure is more or less prototypical, or in other words, whether it is a core member of a category or on the periphery of this caregory. He adopts four criteria: structure, inflection, distribution and frequency. Core members are structurally less marked, show more inflectional distinctions, appear in more contexts and show generally a higher frequency. Iemmolo (2011: 8) adds the criterion of cognitive or semantic complexity: a core member is less complex than its peripheral counterparts.

Morphological marking may also be interpreted in terms of 'iconicity': to make information audible (and visible) by overt morphological marking. 'Economy' is often names as the opponent to iconicity as language structure unfolds between the two poles of *iconicity* and *economy*. The aim of

17

iconicity is to make language (structure) as transparent as possible, whereas economy aims to keep the language (structure) as simple as possible in terms of both quantity and quality. Aissen (2003) as well as de Hoop and Malchukov (2008) offer DOM analyses with iconicity and economy as driving forces. An overview on the two constraints from a functional–typological point of view is given by Iemmolo (2011: 4, 5).

Irrespective of which markedness strategy is followed by any specfic DOM-system, the overt marking correlates with a high degree of animacy and/or definiteness of the DO (referent), while a low(er) degree of animacy and definiteness does not allow for the marking. Between the relatively clear semantic endpoints – i.e. maximally individuated ([human], [definite]) DOs, and not individuated ([inanimate], [indefinite]) DOs – stretches a continuum of 'more or less' animate and definite DOs. The turning point between overt marking and non-marking is subject to cross-linguistic variation: "[Die] Verteilung morphologischer Markiertheit und Unmarkiertheit ist universal, hingegen ist die Lage der jeweiligen Wendepunkte aus den beiden Dimensionen einzelsprachspezifisch" [The distribution of morphological markedness and unmarkedness is universal, while the position of the turning points between the two dimensions are language-specific; translation KAN] (Bossong 1985: 9).

The semantic values may leave open some interpretable range, i.e. 'more or less' definite, specific, or even animate; and so the semantic markedness of DOs is not an absolute but a relative quality. By contrast, the morphological markedness does not leave open any optionality: either there is overt marking, or there is no overt marking.

Most often, DOM-systems do not pattern in such a straightforward way, with one turning point between overt marking and non-marking. If we analyse DOM-patterns, languages with DOM systems have DOs that require overt marking, others reject it, and some others seem to vary. Applied to hierarchies such as given in Table 3 and 4, we have not only one but two turning points. These separate obligatorily marked and unmarked nominals (situated at the edges of the hierarchies), from an optional range; see Figure 1:
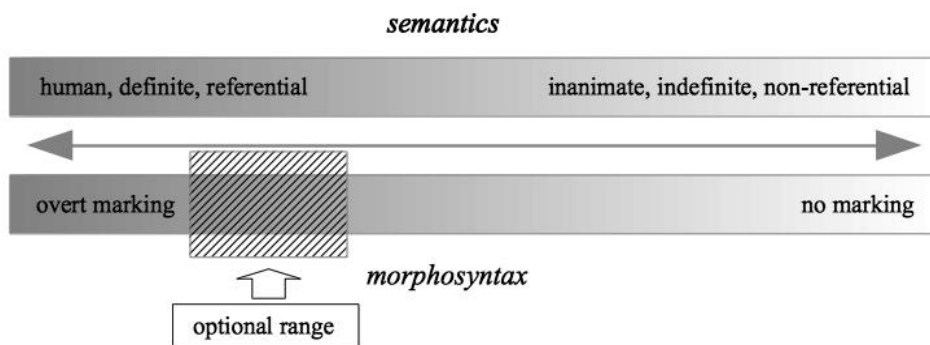
Figure 1: optional marking range (exemplarily), cf. Iemmolo 2011: 50

Optionality shows up in the inconsistent marking of a specific exponent of one specific nominal class, or the inconsistent marking of exponents of certain nominal classes in general. According to Iemmolo (2011: 6), optionality is "synchronic variation" which in turn should "be interpreted as a signal of a diachronic change in action". From a purely synchronic point of view, I argue that optionality in marking is merely relative, and not absolute. We may assume that a speaker consciously chooses a specific structure, because s/he has a specific interpretation in mind. Or, as Croft puts it: "[T]he linguistic structure is shaped by the perspective speakers have of the world" (Croft 2003: 102). Consequently, the features or properties adduced to analyse a given structure are not sufficient nor appropriate to capture and explain the marking pattern. On the other hand, synchronic variation in marking may be due to other causes, such as language contact.

The concepts of markedness and prototypicality, the relationship between semantic and morphological markedness, the relationship between iconicity and economy as general linguistic forces, and finally the locus of optionality, are all of varying relevance for the accounts on DOM that are presented in the following sections. I start with disambiguation, the most traditional of all accounts to explain overt differential marking strategies on the nominal. Subsequently, DOM will be viewed from the perspective of transitivity.

2.2.2 *Disambiguation approaches*

Disambiguation strategies aim for *differentiation*. In the context of DOM, this concerns the morphological distinction between nominals in subject and DO function: if the two nominals are similar or identical in their semantics and formal representation, they must be distinguished morphologically. Hence, disambiguation is required from two different points of view: semantics and

morphosyntax.

These two arguments are discussed by Bossong (1998):

Le marquage différentiel de l'objet est une réponse largement répandue aux *problems syntaxiques* survenues après la perte de la déclinaison casuelle latine, du type indo-européen ancient. Dans le langues romances c'est ou bien la préposition du datif ou bien une préposition spécifique qui sert désormais à marquer les object qui, de par *sémantisme inhérentiel ou référentiel*, ressemblent à des sujets: les objects qui correspondent au prototype objectal ne sont pa affectés pa la margue. (Bossong 1998: 228; emphasis: KAN)

[Differential object marking is a widespread answer to *syntactic problems* occurring after the loss of the Latin system of case declension of the old Indo-European type. In Romance languages the dative preposition or a specific preposition mark these objects, which resemble subjects due to *inherent* or *referential semantics*. Objects which correspond to a prototypical objetc are not affected by the marking; KAN]

NEs in subject and DO function have typical semantic features or values. Subject referents are high in animacy and referentiality; object referents are low (cf., among many others, Comrie 1989; Aissen 2003; Detges 2005). Against the backdrop of markedness theory (as presented above), subjects are semantically and also formally unmarked for these values. Unmarked DOs are low in animacy and referentiality. High degrees of these features are 'unnatural', i.e. marked values for the DO. Overt marking may compensate for this 'defect', which would lead to misinterpretations otherwise. Overt marking allows for the straightforward assignment of the syntactic function of DO. Hence, within transitive structures, DOM offers a disambiguation strategy for certain kinds of nominals. In Romance, that means for full lexical nouns, proper nouns, kinship terms and/or certain proforms.

Disambiguation approaches seem straightforward, but there are many contexts where the disambiguation strategy does not seem a primary goal of DOM. In Romance languages, DOM is not consistently applied when morpological and semantic identity between subject and DO is given, and is applied in contexts where disambiguation is already ensured by other strategies. We find DOM most consistently with certain nominal classes that are already morphologically distinguished by specific forms, for example personal pronouns have different exponenents in subject and DO function, as well as in nominative and accusative case. This holds true for 1st and 2nd person, where DOM often emerges first from a diachronic point of view (cf. Detges 2005: 159), b also 3rd person items. Moreover, these items keep their morphological shape relatively stable: "[P]ronouns usually

display quite a conservative behaviour cross-linguistically, as they tend to retain morphological distinctions that are lost in other kinds of nominals" (Iemmolo 2011: 33). Sometimes DOM actually leads to ambiguity because of morphological or morphosyntactic identity with other nominals in the sentence.

In Romance, the DOM-marker corresponds formally to the dative marker. This holds true for all Romance languages with DOM except for Romanian. In ditransitive constructions, the dative, not following a differential marking strategy, has to be marked obligatorily (cf. von Heusinger & Kaiser 2007: 89). In these contexts, DOM is often suspended, even if the DO would be marked due to its semantics in other contexts (on ditransitive constructions and DOM, see Torrego 1998). Also, locatives marked with the (formally) identical marker may cause ambiguities in combination with DOM-marked toponyms, such as in Corsican or Sardinian. Hence, the morphological disambiguation between the different arguments or nominals, be it in subject, DO or IO function, is no longer ensured.

DOM also appears in contexts where world knowledge allows for the distinction between subject and object, for example with regard to animacy: "it is redundant to mark animacy *per se* as it is clear from the lexical content of the nominal in question" (Malchukov 2008: 210, and others: Moravcsik 1978; Næss 2004: 1188). Additionally, disambiguation may also be provided by word order or intonation. Hence, we may question the motivation of disambiguation from a more general perspective. Beyond these 'phenomenon-internal' issues, and from a more cross-linguistic perspective, it seems that case marking on core arguments is generally absent (cf. Siewierska 1997).

From a diachronic perspective, disambiguation seems not to be the driving force for DOM, since consistent marking systems ensure disambiguation more effectively than differential marking systems. If a consistent system disappears, such as the Latin declension system, the development of a new consistent system would be more logical in terms of differentiation, as has happened for example, for the marking of dative or locative constructions. In his typological study, Sinnemäki (2014) found out that DOM-systems are cross-linguistically and statistically preferred and diachronically more stable than consistent marking systems.

2.2.3 *Transitivity hypothesis*

DOM appears most often in transitive constructions. Since the 1980s, transitivity has been characterised by Hopper and Thompson's (1980) transitivity hypothesis. In their article, they decompose transitivity into components and parameters that contribute to the transitivity of a clause. Transitivity is understood in a traditional sense "as a global property of an entire clause, such that an activity is 'carried-over' or 'transferred' from an agent to a patient" (Hopper & Thompson 1980: 251). The isolated, single NE becomes less important and consequently, the disambiguation between

subject and object is considered of secondary importance (cf. Hopper & Thompson 1980: 259).

Hopper and Thompson's parameters of transitivity focus on participants and the verb, and are listed in Table 6. The verb expresses the activity, and is 'effective'. 'Effectiveness' of the verb concerns the mode of action of the subject and its impact on the object. It is determined by numerous parameters (parameters B–G in Table 6). Hopper and Thompson mention a minimum of two participants (parameter A). The characteristics of the participants are described in the last three parameters (for the subject, parameter H; for the DO, parameters I and J):

Table 6: Parameters of transitivity (Hopper & Thompson 1980: 252)

	high	low
A. Participants	2 or more participants A(gent) and O(bject)	1 participant
B. Kinesis	action	non-action
C. Aspect	telic	atelic
D. Punctuality	punctual	non-punctual
E. Volitionality	volitional	non-volitional
F. Affirmation	affirmative	negative
G. Mode	realis	irrealis
H. Agency	A high in potency	A low in potency
I. Affectedness of O	O totally affected	O not affected
J. Individuation of O	O highly individuated	O non-individuated

In general, the more features from the 'high' column that can be singled out in or attributed to a sentence, the more transitive that sentence is. Transitivity is conceived as a continuum (cf. Hopper & Thompson 1980: 253).

Parameter J, the individuation of the O(bject), has been analysed in more detail. The individuation of the object manifests itself in two different ways. On the one hand it concerns the distinction of the two participants, i.e. the subject from the object. On the other hand, it regards "the extent to which the O[bject] is particularized and viewed as a concrete entity distinct from its background" (Hopper & Thompson 1980: 287). Different features and their values contribute to the concept of individuation (see Table 7).

Table 7: Individuation of the object (cf. Hopper & Thompson 1980: 253; modification, KAN)

	individuated	**non-individuated**
noun type	proper	common
animacy	human, animate	inanimate
concreteness	concrete	abstract
number	singular	plural
countability	count	mass
referentiality	referential, definite	non-referential

As with the composition of definiteness hierarchies or referentiality scales presented in section 2.1, Hopper and Thompson also apply features such as noun type, animacy and referentiality. Beyond these, number, concreteness and countability also contribute to individuation. Yamamoto (1999) adds delimitation and identifiability of an individual entity, which are in turn related to number (i.e. singular, plural). Plurality is considered to weaken the (interpretative) value of animacy because the identity of the referent can be blurred (cf. Yamamoto 1999: 4).

The individuation parameter fulfils a double function: it contributes to transitivity, (see parameter J in Table 6) and it also defines the objecthood or 'object-ness' of nominals: "[…] it seems to us that the tendency to mark just definite/animate O's reflect the purer object-ness of such O's, and simultaneously marks the higher Transitivity of the clause as a whole" (Hopper & Thompson 1980: 291). In contrast to Croft (1988) and Comrie (1989), who state that typical DOs are low in animacy and definiteness, Hopper and Thompson (1980) consider definite and animate DOs to be typical objects. This is supported structurally; cross-linguistically, indefinite and inanimate nouns are found in intransitive or even incorporation structures (cf. Hopper & Thompson 1980; Næss 2004; Iemmolo 2011): "As noted by Næss (2004: 1192), the idea that direct objects low in individuation are the most typical ones could be highly problematic, since these clauses do not involve a direct object at all from a structural point of view" (Iemmolo 2011: 30).

Assuming a high degree of individuation in terms of specificity or definiteness as a requirement for (direct) objecthood can also be problematic if we consider indefinite or negative quantifiers. Corsican *nimu* or Spanish *nadie* 'nobody', as well as Corsican *qualchidunu* or Spanish *alguien* 'somebody', are mostly *a*-marked, but no specific reading is available for these items. Hence, the relevance of specificity is not unequivocal for DOM (cf. Rodríguez Mondoñedo 2007; López 2012).

Against the theoretical background of transitivity, Hopper and Thompson (1980) investigate DOM. Only objects which are highly individualised get DOM-marked; in particular, referential objects with animated or human referents. The extent to which these different features contribute to the individuation of the object may vary from language to language.

Despite the persistent importance of Hopper and Thompson's (1980) account, some points must be looked at critically. The absence of hierarchisation in Table 6 suggests all parameters are of equivalent significance in defining transitivity. However, while some of these parameters are essential (such as the number of participants: two distinct participants are indispensable), other properties are ranked differently in different language varieties (such as agentivity and volitionality of the agent). Also, the mechanism of co-variation between the parameters seems more complex than that proposed by Hopper and Thompson (1980). Co-variation among *all* parameters does not seem reasonable. Lastly, the way Hopper and Thompson relate markedness to prototypicality contradicts traditional conceptions of markedness, as presented in section 2.2.1 above. See also Iemmolo (2011: 16) for a short summary of the critique of Hopper and Thompson's account (1980).

The parameters of Hopper and Thompson (1980) are adopted by other authors working on DOM. Of the two object parameters – affectedness and individuation of O – Næss (2004) prioritises affectedness (see Table 6, parameter I):

> The tendency to case-mark objects that are high in definiteness and animacy is in fact a reflection of the accusative case as marking objects which are construed as being highly affected […] definiteness and animacy provide the yard sticks by means of which affectedness may be measured. (Næss 2004: 1203)

Næss assumes that affectedness is the principal property of DOs, to which individuation contributes: if the object exhibits a low degree of individuation, it is rather unlikely that the respective nominal can be encoded in an argument position, and instead it alludes to intransitive constructions or incorporation structures where nominals may not fulfil a DO function (cf. Næss 2004: 1192).

Beyond individuation, three more aspects contribute to Næss' (2004) affectedness concept: (i) quantification, (ii) referentiality and (iii) salience. With regard to (i), objects which are affected as a whole by the action expressed by the verb, are more affected than objects that are only affected in parts. This distinction can be expressed lexically by a determiner or definite article for complete affectedness, or by certain quantifiers, such as *some* or *many*, to show partial affectedness. With regard to (ii), the determiner is related to a high degree of referentiality. By way of illustration, recall the exemplification of specificity and definiteness by von Heusinger and Kaiser (2003) and Hopper

and Thompson (1980) given in section 2.1, with Examples (1) and (3a and 3b). Highly referential objects are considered highly affected by Næss (2004). And finally, with regard to (iii), salience is related to animacy. From an anthropocentric point of view, animated or human referents are of greater importance than inanimate ones: "[…] effects on human or animate entities are perceived as more dramatic, more significant, than effects on inanimates […]" (Næss 2004: 1202).[5] 'Affectedness features' are already included, explicitly (animacy, referentiality) and implicitly (countability), in Hopper and Thompson's (1980) individuation concept (see Table 7, above).

If we summarise the semantic approaches, we may conclude that disambiguation focuses on the semantic and morphological distinctiveness of the two nominals in subject and object function. This holds true for all the accounts presented, independent from their theoretical background. The transitivity accounts are all concerned with the transitivity of whole constructions, including subject, verb, and DO. They put single parameters to the fore; for the DO this parameter is either individuation (Hopper & Thompson 1980) or affectedness (Næss 2004).

2.3   Syntactic accounts on Romance DOM

Syntactic accounts on Romance DOM deal with questions of case assignment, such as how to distinguish (or not) between dative and accusative case and the structure of the DP. They follow one of two basic premises: either *a*-marked accusatives and *a*-marked datives are structurally the same or at least similar, while unmarked accusative objects are different; or dative and accusative are different, and all accusatives, either *a*-marked or not, are the same. A third and different strand is offered by accounts that place the phrase structure of the DP at the centre of investigation. The idea that the *a*-marker and the definite article share properties and functions in terms of Case assignment builds on accounts of split DP. This idea has been elaborated for Romance varieties by Giusti since the 1990s, and is further elaborated for Corsican DOM, in this thesis.

Among the adherents of the first line of thinking are such authors as Torrego (1998), Rodríguez Mondoñedo (2007) and Ormazabal and Romero (2013): they assume that the *a*-marked accusative object and the dative object are related semantically, morphologically, and also syntactically. Two nominals are inserted in the very same syntactic position, or at least *a*-marking for accusative is the result of processes similar to dative marking. Brugè and Brugger (1994) and López (2012) instead keep dative and accusative case separated. They assume a different syntactic process and a distinct functional phrase (FP, KP) hosting the DO. These accounts differ in their realisation of the *a*-marker, whether in syntax or post-syntactically, by different spell-out or vocabulary insertion rules. For all

---

[5]   For more on salience, in relation to definiteness, see von Heusinger (1997).

these syntactic accounts (save Brugè & Brugger 1994) it holds true that all *a*-marked DOs need to raise or be raised out of the complementiser position of the verb phrase, CompV, into a higher position. This can be Spec*v*P – the specifier of the verb phrase (*v*P) (e.g. Torrego 1998), a dative phrase (DatP) above *v*P (cf. Rodríguez Mondoñedo 2007), or a position lower than *v*P that might be the specifier of a functional category *α*, such as Spec*α*P (e.g. López 2012).

The third strand of syntactic accounts elaborate the phrase structure of DPs in detail, based on accounts of split DP. I consider the accounts of the following three authors, Zamparelli (2000), Giusti (2015), and Wiltschko (2014). Zamparelli and Giusti focus on the DP-structure and functions in more traditional terms: Zamparelli's idea is driven by functional distinctions of different DP-layers regarding, for example, type-shifting; while Giusti emphasises the similarities of functions within the DP, for example, case marking and definiteness marking. These affinities between case, definiteness and the projection of a determiner position or phase, i.e. the D-projection also link Giusti's and Wiltschko's accounts: Wiltschko broadens the view beyond nouns and noun-related functions, and presents universal functions and categories.

Before going into the (morpho-)syntactic details of these accounts, let us start with some basic morphological remarks on Romance DOM: the accusative noun in DO-function in *a*-marking DOM-languages is either syncretic with the nominative when unmarked, or syncretic with the dative (or locative) when *a*-marked. Of the two accusative objects, it has been the *a*-marked one that has always been of major interest for the development of syntactic accounts on Romance DOM. Among the unmarked DOs only specific classes or constructions have attracted major attention, e.g. bare nouns.

Regarding the (non-)persistence of the *a*-marker, it is noteworthy that DOM-*a* often disappears in ditransitive constructions (see Torrego 1998). In approaches where DOM is related to dative in structural terms, this is explained by the assumption that both objects compete for the same case checking or case assigning position. DOM-objects can also be syncretic with locatives marked by *a*. Unlike ditransitive contexts, DOM-*a* is never suspended in these contexts (cf. Fábregas 2013: 32). Beyond that, not all Romance DOM systems show the accusative–dative syncretism: Romanian DOM is expressed by *pe*, which may appear in locative contexts, but is never used to mark dative.

I will focus on the nature of the *a*-marker and its realisation: Is *a* the exponent of a functional category heading a phrase of its own, representing a maximal projection? Or is it located in a specifier position? Does *a* contain (a bundle of) feature values, e.g. [specific], to contribute to the semantic interpretation of the sentence? Are there different (homophonous) *a*-markers within one language? In some cases the *a*-marker seems to bear some meaning contributing to the interpretation of the

nominal, such as [specific]. In other contexts, and/or with regard to other features, *a* does not add any (semantic) information, e.g. animacy.

Attempts to answer these questions have been made by the authors whose accounts I present below. To understand better the internal structure of DPs, I start with the accounts on split DP in section 2.3.1, before coming to the so-called 'dative accounts' on DOM in section 2.3.2, followed by the 'accusative accounts' on DOM (section 2.3.3).

2.3.1   *The DP-structure: splits of different kinds*

Split accounts have started with splitting up the so-called functional categories, namely: inflection (INFL), including tense (TENSE) and agreement (AGR), cf. Pollock 1989; or complement(ation) i.e. complementiser phrase (CompP or CP), including different phases such as ForceP, FocusP, TopicP, FinP, cf. Rizzi 1997).[6]

For the determiner (D), we can identify (at least) two main strands to motivate a split DP-hypothesis: on the one hand to explain the structural similarities between different phrases such as the inflection phase (INFLP or IP) and the determiner phrase (DP), the distinction between functional and thematic phrases, and the movement of elements within phrases, such as the movement of the verb into the INFL or TENSE phrase (V-to-INFL/TENSE) and the movement of the noun into the DP (N-to-DP). On the other hand, to account for the semantic distinction between different determiner elements and the need to host elements or features that contribute to the semantics commonly attributed to the DP level, such as e.g. reference or identification.

Comparing the sentence structure, [$_{CP}$ [$_{INFL/TENSEP}$ [$_{VP}$ […]]]], with the DP-structure, we can identify three domains: the topmost domain is the area of discourse relevant information, in the middle is the morphological/inflectional area, and at the bottom is a thematic area where the lexical head is projected. In functional terms, D serves some of the same functions as INFL. While INFL is about time reference, D relates an individual to an utterance. In terms of definiteness, D "relates the referent to the discourse by marking whether it serves as novel or a familiar discourse referent" (Wiltschko 2014: 77, referring to Heim 1988).

To fulfil these (and other) functions, the functional head D has different properties. For example, it may host the definite article or a case assigner, it is a locus of the semantic notion of definiteness, and it grants argument status to the nominal by turning a predicate into a referring expressions. With regard to this last, there are controversial issues. Longobardi (1994) elaborates a uniform structure for different kinds of nominals in argument position, to show that the DP-structure

---

[6]   Splitting up the structure into functional phrases, where each feature corresponds to one projection leads to the cartographic approach (Cinque 1999, and subsequent work).

is required in order for a nominal to function as an argument. By contrast Chierchia (1998) states that there is no universal DP category for argument nominals. He refers back to Carlson's (1977) idea that nominal phases (NPs) can denote predicates and kinds and can also be referential expressions. The interpretation varies due to type-shifters that are not necessarily determiners or articles. Giusti (2015) offers a detailed description of the weaknesses of both accounts, also connected to a broader theoretical and empirical background (for Longobardi 1994, see Giusti 2015: 63, 64; for Chierchia 1998 and his Carlsonian approach, see Giusti 2015: 66, 67).

Along similar lines, Déchaine and Wiltschko (2002) also argue that a DP-layer is not a necessary requirement for argumenthood. They investigate pronouns, their internal structure and syntactic behaviour. According to them, the nominal class 'pronoun' is not a primitive (cf. Déchaine & Wiltschko 2002: 409), but proforms are compositional, i.e. they build up a layered structures, a so-called DP-shell or only parts of it. Déchaine and Wiltschko distinguish between pro-DPs, pro-φPs (where φP represents the φ phrase) and pro-NPs: "The categorical status of these pronominal categories determines their external syntax and their inherent semantics, which in turn determines their binding-theoretic status" (Déchaine & Wiltschko 2002: 410). For the internal structure of the pro-phrases, see the following structure (1):

(1)   **Structures of pro-phrases** (cf. Déchaine & Wiltschko 2002)
- pro-DP: [$_{DP}$ [D][$_{φP}$ [$φ$][$_{NP}$ [N]]]]
- pro-$φ$P: [$_{φP}$ [$φ$][$_{NP}$ [N]]]
- pro-NP: [$_{NP}$ [N]]

Thus, the two subconstituents of DP, φP and NP, can function each as an independent proform. Pro-φPs can function as arguments; they have no internal restriction on their distribution or function (Déchaine & Wiltschko 2002: 410, examples p. 416).

In general, it is rather difficult to distinguish the concept of D as a syntactic position or category from the linguistic item(s) realising it. In many languages, the definite article realises D.

The element 'definite article' has been approached in different ways: as an item which expresses the semantic notion of definiteness (cf. Lyons 1999); a "subordinator assigning argumental status to its NP complement" (Alexiadou et al. 2007: 55, following Abney 1987, Stowell 1989 and Szabolcsi 1994); the element that bears referentiality (cf. Loebel 1989, Longobardi 1994); or, contrary to the semantic approach of Lyons (1999), a grammatical morpheme primarily assigning

case to its complement (Giusti 1993, 1997 and subsequent work). In some languages, articles are inserted on syntactic grounds regardless of any referential or semantic properties of the noun phrase, i.e. the definite article as a grammatical morpheme, deprived from semantic or descriptive content. Syntactically definite articles behave differently from other determiners: only definite articles are considered to be located in functional heads; other determiners, e.g. demonstratives, are phrases and may appear prenominally or pronominally, marked with [pre] or [pro] respectively in the following example 4:

(4)   a. *I finally bought [this]$_{pre}$ book Lisa is always talking about.*

         b. *Lisa is talking about [a book] since weeks. [This]$_{pro}$ is what I bought today.*

They imply referentiality in terms of deixis. Generally, they cannot refer to kinds, and no generic reading is available (cf. Alexiadou et al. 2007: 76, 95, 99).

Before splitting up the DP into different projections with more than one D-head, the early DP accounts are characterised by a single D-head that hosts simultaneously "features of (in)definiteness, gender, number and, crucially, Case" (Alexiadou et al. 2007: 152). Case then, was the first of these features to be split from the D-head: Lamontagne and Travis (1987) introduce an additional functional layer above DP, the *Kase Phrase* (KP). Loebel (1994) also distinguishes between K (for case) and other functions of D; she splits "the category D into two separate functional categories, one with semantic content, where D itself functions as a bearer of referential features, and one with primarily syntactic functions, i.e. K for Case" (Loebel 1994: 51). For Loebel, the syntactic function of case interacts with the referential features of D: [± definite] and [± specific]. But it also bears features of the noun and the nominal's referent in particular: [± human] and [± animate].

Since the mid-1990s Giusti advocates for a unified account of case morphemes and definite articles. In Giusti (1995b), she suggests a unifying functional head, which both fulfils case requirements and saturates the predicate nominal, i.e. turns the predicate into an argument: "[…] it can be argued that the DP-projection, and in particular its property of being the 'saturator' of the predicate NP (cf. Higginbotham 1983; Longobardi 1994), can be conflated into the same functional head as Case/case" (Giusti 1995b: 77). This common functional head allows a universal structure for nominal phases in all languages, irrespective of the presence of overt morphological case and/or a definite/indefinite article. This head combines both the functions of a KP and a DP.

Beyond case, other functions are also associated with D, for example deixis and determination.

If we split the DP according to these two concepts into a simplified structure of the type [DP1 [DP2 [...]]], each DP can be associated with one function, one deixis, and one determination. According to Alexiadou et al. (2007), the topmost DP1 encodes discourse and pragmatic aspects of the DP's interpretation (referentiality, deixis, familiarity) and the lower DP2 encodes determination, i.e. (in)definiteness. Wiltschko (2014) interprets deixis (and anaphoricity) as the core manifestations of the anchoring function of the nominal that relates the nominal to the utterance; also, case expresses the anchoring function (for more detail, see further on).

In the following paragraph(s), I talk about Zamparelli's (2000) Layered Determiner System, Giusti's (2015) account on the case-expressing definite article, and Wiltschko (2014) and her Universal Spine Hypothesis. I have selected these accouns for several reasons. It is common to all accounts to split up the functional layers within the DP and to question the traditional concepts of categories, word classes and related functions. This allows for (new) allocations of elements, processes, functions and features within the DP, namely determination, individuation and case, which are among the most important concepts relating to DOM.

Zamparelli (2000) is concerned with the syntactic structure of DPs, the functional layers and the distribution of linguistic elements and their interpretation. He focuses on the 'determiner system': the left edge of the DP, i.e. the functional layers 'above' a lexical core (i.e., NP). His split DP is motivated by different phenomena, such as the distinction between a predicative and an argument status of noun phrases, the distribution of features and properties commonly attributed to the DP level, and processes taking place at the DP level, such as type shifting.

Zamparelli's determiner system contains two phrases: uppermost is the Strong Determiner Phrase (SDP), followed by the Predicative Determiner Phrase (PDP). Below, in the structure (2), is the Kind Determiner Phrase (KIP) not specified in detail by Zamparelli and not of major importance to my approach.[7]

(2)   **Determiner system**: SDP_e [$PDP_{<e,t>}$ [KIP_e [... [NP]]]] (Zamparelli 2000: 16, Fig. 25)

Looking at the structure in (St. 2), we find PDP below SDP. According to Zamparelli, every argument noun phrase contains a hidden predication: the PDP. The PDP is the noun phrase's internal predicate and expresses a 'timeless' property; it can be negated and modified and is purely property-denoting.

---

[7]   As for the KIP layer, Zamparelli admits that its structure and semantics are rather sketchy. He assumes that nouns are not inserted as properties – hence, not as elements of type $<e,t>$ – but that property-denotation is one of the functions of the PDP-layer (cf. Zamparelli 2000: 17).

It never appears in argument position (cf. Zamparelli 2000: 16). The denotation of the noun phrase depends on the internal structure and the position of the nominal within the DP, rather than the nominal's position within the sentences (cf. Diesing 1992): in the highest projection (i.e. SDP), the nominal denotes an individual, while within the PDP it denotes a property.

In terms of semantic types, the predicative noun phrase (i.e. PDP) is of type $<e,t>$, while the argument, referential or quantificational noun phrase (i.e. SDP) is of type $<e>$. SDP as the topmost layer is the only locus for an interpretation of semantic type $<e>$. The $<e>$ type noun phrases may be further distinguished according to whether or not they need to undergo Quantifier Raising (QR) to be interpretable.[8] While for Wiltschko (2014) interpretation and multifunctionality of linguistic elements is due to underspecification of values and/or different association-processes (more details further on), for Zamparelli (2000), the nominal's interpretation is influenced by three factors: the presence of the SDP-layer; the overt realisation of SDP; and QR, or Quantifier Construal (QC), of the SDP-nominal (for more on the specificity of these processes, see Zamparelli 2000: 4).

If we apply Zamparelli's split DP-structure to pronouns, for example, taking into account the different functions they may fulfil, we get three different types of pronoun, as shown below (3):

(3) **Structural differences of Italian pronouns** (Zamparelli 2000: 19)

SDP   *lo*$_{+Agr}$ pronominalises full argument noun phrases

PDP   *lo*$_{-Agr}$ pronominalises predicate nominals, also predicative adjectives

KIP   *ne* pronominalises a subpart of a full noun phase, including the noun and some adjectives, but not quantifiers

The syntactic structure of the determiner system presented in (St. 3) has an impact on the denotation and interpretation of the lexical and functional elements within it. Determiners, a term that, according to Zamparelli, includes articles, demonstratives and quantifiers, may be interpreted differently "depending on whether they are located and interpreted in SDP […], or in PDP". That is, the variation of the determiner depends on and corresponds to structural differences (cf. Zamparelli 2000: 17).

With regard to the presence or absence of multiple determiners, Zamparelli's split account offers positions to host more than one determiner: at least one in each of SDP and PDP. The lower

---

8   Partee (1987) proposes three types of noun phrase – referential, predicative, and quantificational – with corresponding types of denotations – e (individuals), $<e,t>$ (set of individuals or properties), and $<<e,t>t>$ (set of properties), respectively. The denotation may change if the language provides the respective means/operators (cf. Zamparelli 2000: 7).

one, PD, can be assimilated to a pre-N adjectival position. This fits well to explain numerals, after a definite article at this position, as in, for example, *the two boys…*. In many cases two determiners may not appear simultaneously: [[**the a*] *man*] or [[**two some*] *people*]. Here, Zamparelli argues in semantic terms: two determiners with overlapping meaning cause redundancy and are therefore excluded:

(4) **Redundancy**: Two functional words F_i, F_j within the same DP give an impossible representation if the meaning of F_i entails the meaning of F_j or vice versa. (Zamparelli 2000: 131 (333))

Two determiners within one DP are possible only when each one adds something to the meaning of the DP. With reference to the example above, *the two boys*, *the* adds definiteness, and *two* cardinality.

Zamparelli (2000) assumes D to be a variable (cf. Zamparelli 2000: 162). (Strong) determiners which appear in SDP are presuppositional: a filled SD^{max} level triggers a presupposition of existence for the objects over which the variable ranges. As Zamparelli says, "In my system [see (5)], this [the presupposition of existence] can be cast as a property of a syntactic layer, rather than a property of a special reading of a class of words".

(5) **SD^{max} Presupposition**: A filled SP^{max} layer triggers presuppositions of existence. (Zamparelli 2000: 127 (321))

As proper names, and pronouns too, appear at the SDP-level, they also trigger a presupposition of existence, referred to by Zamparelli as the Descriptive Content Condition:

> Proper names, common nouns and pronouns in their referential usage are anchored to specific individual entities. Something unifies all the strong cases: the Descriptive Content Condition. As noted, with strong determiners, referential pronouns, proper names and maybe bare plurals, there exists a presupposition that the set SD ranges over is non-empty in the current world. (Zamparelli 2000: 163)

To summarise, Zamparelli's determiner system contains two relevant layers, SDP and PDP. SDP denotes an individual, PDP a property. Moreover, SDP is related to the argument status of nominal; it contains nominals of the semantic type $<e>$ and (strong) determiners in SDP trigger a

presupposition of existence. Zamparelli follows an established assumption about the DP-structure: the topmost-DP level's requirement for a referential reading, and also for the argument status of the respective nominal. By offering two DP-layers, he can account for different interpretations of DPs (e.g. with regard to referentiality or genericity) and the (im)possibility of having multiple determiners.

Contrary to Zamparelli (2002), Giusti (1993, 1995a, 1995b and subsequent work) maintains the early idea of a single functional category D (in earlier work F), and subsumes case and reference "under a single functional head" (Alexiadou et al. 2007: 156). For her, the definite article, realised in D°, is case rather than the realisation of referentiality.

Giusti's (2015) account offers a minimalistic approach to the left edge of the nominal domain (similar to Zamparelli's 2000 determiner system), the complementation layer. Giusti applies Rizzi's (1997) account and terminology: Rizzi (1997) proposes a tripartite structure of the nominal parallel to the clause structure. The structure is organised in the following way: complementation layer > inflection layer > lexical layer; whereby the complementation layer is the so-called 'left edge'. The complementation layer refers to the extended projection of the lexical head and is the interface with the discourse. Giusti's idea is about one single functional projection, DP, where D° exclusively hosts the article. The article, in turn, is a discontinuous part of N, i.e. nominal morphology, namely case.

Giusti consistently differentiates between *determiners* and *articles.* Determiners build up a class containing different kinds of items, such as demonstratives, possessives and others, while articles are explicitly not among these items. Different from traditional accounts on DP, Giusti's definite article does not carry reference. Reference is provided by a referential index ι, which is 'outsourced' to a specifier position of the DP, SpecDP, and which may (but does not necessarily) co-occur with the definite article: "This predicts that the grammatical element *the* does not always appear with indexical interpretation and vice versa that we can have indexical interpretation even without *the*, under particular syntactic conditions" (Giusti 2015: 56). Detaching reference from the article or, in structural terms, splitting the article and the referential index ι, has several advantages: on the one hand, the presence of the article does not force indexical interpretation, but allows for generic interpretations. On the other hand, an independent referential index allows for other items than articles, e.g. personal pronouns, demonstratives, and proper nouns, which bear such an index inherently.

Additionally, Giusti (2015) assumes that the [person] feature is not part of the feature bundle of N. As for the referential index ι, [person] is also located in the left edge of the nominal domain

where both can merge "with N as the highest specifier, the Left Edge" (Giusti 2015: 50). Similarly, Wiltschko (2014: 11) associates [person] with indexicality under the condition that [person] must be located in the left edge of the DP. In their account, [person] may also be interpreted in a non-indexical way when it is located in their so-called φP, the phrase below DP (see proform-structures in (2)).

The referential index ι, sometimes also called the ι-operator, makes the DP an individual constant – parallel to other determiners, such as demonstratives, personal pronouns, and proper names (cf. Giusti 2015: 56). These elements may all occupy the left edge of the NE, i.e. SpecDP. See the following structure in Figure 2:

```
                    DP
                  /    \
              SpecDP    D'
            ⎧ ι-operator ⎫  /  \
            ⎨ demonstrative ⎬ D°   ...
            ⎪ pronoun      ⎪ (definite
            ⎩ proper name  ⎭  article)
```
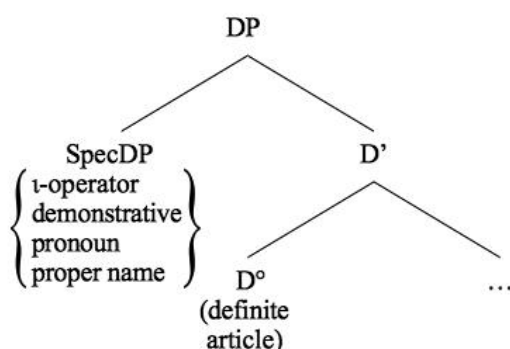
Figure 2: Left edge of the nominal element: SpecDP (Giusti 2015: 56)

Giusti identifies the specifier position of DP as relevant for the interpretation as an individual, by hosting the ι-operator. Zamparelli (2000) attributes this function to the SDP-layer in general, i.e. the realisation of D at the SDP level. Common to both accounts is the allocation of individuation at the topmost layer, the left edge of DP or the leftmost of multiple DPs.

For both Wiltschko (2014) and Giusti (2015) case and the functions associated with D/D° are closely correlated. The terminology differs as far as Wiltschko describes case as an anchoring strategy in the nominal context parallel to other deictic and anaphoric functions and elements. Giusti (2015) applies more traditional (generative) terms, speaking about the left edge of D and the accessibility of the NE to the next phase. Even if Wiltschko's head positions also bear unvalued features ([$u$ident] (unvalued identity feature) and [$u$coin] (unvalued coincidence), in the nominal and verbal domains, respectively), she is much less concerned about single features than Giusti. Giusti follows a feature-oriented approach, in so far as interpretation and movement are triggered and influenced by it. She assumes heads to be hierarchically ordered feature bundles, invoking Matushansky (2006): "A head is a syntactically indivisible bundle of formal features" (Matushansky

2006: 70). This hierarchy of features is given by Universal Grammear (UG), as presented by Giorgi and Pianesi (1997).[9]

A head merges as a complete feature bundle that is constituted pre-syntactically, in a structured way.[10] The bundle is the result of morphological processes: "The bundle itself is not created in the syntax, at the cost of extending the application of Merge to subparts of terminal nodes (*pace* nano-syntactic approaches)" (Giusti 2015: 118). On the other hand, the bundles are (already) part of the lexicon: They are constituting part of paradigms, which, in turn, structure the lexicon.[11] Any split approaches (e.g. split DPs) reveal (information about) the internal structure through the order of the extended projection(s).

In this vein, Giusti (2015) proposes a hierarchically ordered feature bundle for the nominal which directs the process to extend the projection, as shown in the following structure (6):

(6)   **Hierarchy of the nominal features**: Case (= uPred) > Number > Gender > N ι

How the hierarchy applies in syntax depends on the single item and its paradigm:

> How the bundle is realized on each segment (for example if Case is realized as a preposition or as a scattered morpheme of N in the shape of the article, or as a morpheme on N, or a combination of these) will depend on the inflectional paradigm of individual vocabulary items. (Giusti 2015: 117)

A feature based approach also requires feature sharing processes. For Giusti "not all instances of feature sharing are instances of Agreement in the sense of Chomsky (1995)" (Giusti 2015: 80). She conceives feature sharing as the result of different processes, which all require 'Merge' (cf. Giusti 2015: 83). Merge, in general terms, satisfies 'Selection' and 'Modification' (for details, see Giusti 2015: 83, 84) and concerns three processes, 'Agreement', 'Concord' and 'Projection'. Regarding the

---

[9] Giorgi and Pianesi (1997: 14–16) assume that functional features are ordered hierarchically according the Universal Ordering Constraint: features are ordered so that, given F1 > F2, the checking of F1 precedes the checking of F2.

[10] Giusti (2015) defends a somewhat lexicalist view: the elements to be merged into the derivation bear features valued pre-syntactically. While she supports a rich syntactic structure, similar to cartographic approaches, Giusti assumes that bundled features merge directly. This avoids or at least minimises the creation of (vacuous) structure. Giusti allows for the same feature to remerge.

[11] The idea of heads as a feature bundle forces us to shift between syntax and morphology. For syntax, heads are 'one item', a feature bundle whose internal structure is not accessible to syntactic processes. From a morphological point of view, we may investigate how these features are bundled pre-syntactically or think about how a potential feature hierarchy influences morphological and syntactic processes. For more on the topic, see e.g. Matushansky (2006) and her idea about syntactic Merge and a morphological merge, i.e. 'm-merger', working at the interface between morphology and syntax.

differences between these, Giusti states that:

1. 'Agreement' refers to a head argument relationship: the head selects the argument. Agreement results in Case assignment. It is a relationship that involves a probe and a goal. The nature of the relationship is symmetrical, as agreement satisfies (un)interpretable features for both items. The relevant feature involved here is [person], one of the φ-features.

2. 'Concord' refers to the relationship between an extended projection (phrase-type) and a head: the extended projection modifies the head – not further relevant for my argumentation here.

3. 'Projection' is concerned with a remerging head that builds up the extended projection. The head is bundled with interpretable and uninterpretable features. The remerging and feature sharing is due to the scattered/split realisation of the features in the bundle (cf. Giusti 2015: 85).

The differentiation between different kinds of feature sharing – i.e. Agreement, Concord, and Projection – allows Giusti to label only one specific process, the head argument relationship, namely Agreement. While Projection is interesting from a general point of view as it concerns the ordering and distribution of nominal features, Giusti's Agreement is interesting as it deals with [person] and Case assignment. It also explains the evoked parallelism between Agreement and Binding: Giusti assumes that Agreement and Binding relate to the same features, namely [person], following Watanabe (2000):

> Binding clearly regards the identification of a referent in the world or in the discourse. It is therefore related to a referential index [ind]. The minimal elements that carry this index are personal pronouns. For this reason […] I proposed that the feature involved in Agreement is Person. (Giusti 2015: 95).

To sum up, Giusti splits the DP according the feature bundle commonly associated with D and/or N. Neither is [person] part of the nominal features bundled in N, nor the referential index ι part of D°. Instead [person] and the referential index ι merge with N as the highest specifier, the left edge. Being "the only position transparent to the next phase, therefore transparent to the probe of the next phase" (Giusti 2015: 50), the left edge is the interface with, or connector to the next piece of discourse. Against this backdrop is seems most logical to relate Case to this position (cf. Giusti 2015: 119).

Wiltschko's (2014) account broadens the view; she is concerned with the universality of categories in terms of functions. Her *Universal Spine Hypothesis* emphasises the functional (and sentential)

similarities between the verbal and nominal domain(s). The structure itself is category-neutral (i.e. in its core, the same for the verbal and nominal domain(s)). The spine "defines the core functions associated with each domain" (Wiltschko 2014: 22).

She assumes a set of universal categories C_{UG}, where each universal category K represents one specific functional domain. The functional domains are hierarchically ordered, as shown in (7)).

(7)   **Set of universal categories C_{UG}** = universal category K_1 > universal category K_2 > universal category K_3 > …

The function of each K is relational: "For example, the anchoring category (K:*anchoring*) relates the event to the utterance" (Wiltschko 2014: 28). This is reflected by the internal structure of the categories. Wiltschko (2014) assumes the same internal hierarchical organisation for these functional categories as for syntactic constituents, namely organisation by c-command, as shown in (8):

(8)   **Internal organisation of K:** [$_K$P arg [$_K$ K, arg]]

Like constituents, K is endocentric ("K and its complement construct a complex K, which bears the same label as K") and extendable ("A complex K may be extended to include another UoL [Unity of Language] without changing the label of the newly formed expression") (Wiltschko 2014: 60).

Wiltschko (2014) also assumes valuation processes in syntax: Any head K bears an unvalued feature that must be valued by a valuation process. Wiltschko distinguishes *three* such processes. First, morphological marking, which is (somewhat) independent of syntax; Wiltschko calls it internal valuation, or m-valuation (i.e. morphological valuation). Then, two types of external valuation. Then, the clause headed by the respective K head is embedded either by a matrix predicate, pred-complementation, or by a higher functional head, i.e. f-complementation (cf. Wiltschko 2014: 145, 146). This results in pred-valuation or f-valuation. The unvalued feature in the context of nominals is [$u$ident].

In addition to these Ks we have language-specific categories c, constituting a set C_{Lg}, as shown in the following (9).

(9)   **Set of language-specific categories C_{Lg}** = language-specific category c_1 + language-specific category c_2 + language-specific category c_3 ...
language-specific category c = *universal category K + linguistic element*
$c \neq K$

These categories are built by matching a language-specific linguistic element with the universal categories K. In other words, $c = K$ + linguistic element. Hence, K is not only a static device in terms of a universal category, but interacts also with language-specific elements to build up the language-specific categories C_{Lg}. Wiltschko emphasises the importance of distinguishing between the universal category K and language-specific category c, i.e. $c \neq K$. As the language-specific categories C_{Lg} are constructed on the basis of varying language-specific linguistic elements, C_{Lg} also vary. The interpretation of linguistic elements depends on their syntactic context. This multifunctionality

> lies in the recognition that categorical identity mediates between a UoL, and its interpretation. The categorical identity of a language-specific category is not just a matter of the UoL. Instead, the UoL is viewed as an ingredient in the construction of the category with the universal categorizer K. Consequently, the interpretative effect that correlates with a switch in categorical identity of a given UoL must be due to K. (Wiltschko 2014: 26)

To summarise, the UoL associates with K to get c. To account for the diversity of language-specific categories and language-specific linguistic items, it is important *how, when* and *where* these linguistic items associate with K (for details, see Wiltschko 2014, Chapter 3).

Four functional domains, represented by single Ks respectively, build up the Universal Spine: 'discourse linking', 'anchoring', 'point-of-view' and 'classification', ordered in relation to each other in the hierarchy given in Table 8. Wiltschko assumes that lexical categories are dominated by several layers of functional categories. For example, in the domain K:*classification*, these would be discourse linking, anchoring, and point-of-view.

Table 8: Functional domains in the Universal Spine (cf. Wiltschko 2014: 28)

The Universal Spine C_{UG}: *K:discourse linking < K:anchoring < K:point-of-view < K:classification*
• *K:discourse linking*: relationship between the proposition/referent and the ongoing discourse • *K:anchoring:* anchoring the event/individual to the utterance; deictic and anaphoric anchoring • *K:point-of-view:* viewpoint relative to which the event/individual is presented • *K:classification:* event or property/individual?

The motivation for these domains derives from *v*P-shell and split-INFL approaches, hence the verbal domain. Traditionally, "the functional structure above the VP can be divided into several distinct areas […] each area is associated with a particular set of roles that the nominal arguments introduced

in the VP may bear": thematic roles, grammatical roles, and discourse roles. Sometimes referred to as theta positions, A[rgument]-positions and A'-positions and are closely related to the head positions in the corresponding area, little *v,* INFL (or T), and complementiser C (cf. Wiltschko 2014: 72). Wiltschko extends the tripartite structure for a forth domain and in doing so she returns to the INFL-split, introduced by Pollock (1989): her point-of-view refers to Aspect, and the anchoring domain refers (among others) to Tense.

From the functional domains presented in Table 8, κ:*anchoring* is of major interest for this thesis. Due to its category-neutrality, κ:*anchoring* may appear in the nominal or verbal context, where different concepts or features may fulfil the anchoring function. Core manifestations of anchoring, with respect to nominals, are deixis and anaphoricity. (In the verbal domain, tense, but also person, aspect or location, may anchor the event to the utterance (see Wiltschko 2014, Chapter 4)). Among nominal anchoring elements, Wiltschko investigates demonstratives, pronouns and case. To derive deictic and anaphoric categories, Wiltschko (2014) mentions two main ingredients: first, the abstract argument arg_{ind} that is the complement of the head of κ:*anchoring.* It is intrinsically anaphoric and refers to an individual. And second, the intrinsic, unvalued feature associated with the head of κ:*anchoring,* [$u$ident]. This unvalued *identity* feature is the defining feature of nominality: "Nouns carry, as part of the meaning, criteria of identity […] These are criteria that determine whether the individual denoted by one instance of a noun is the same as the individual denoted by another instance of that noun" (Wiltschko 2014: 208, 209; citing Geach 1962).

Shown in a syntactic tree, nominal anchoring is structured in the following way:
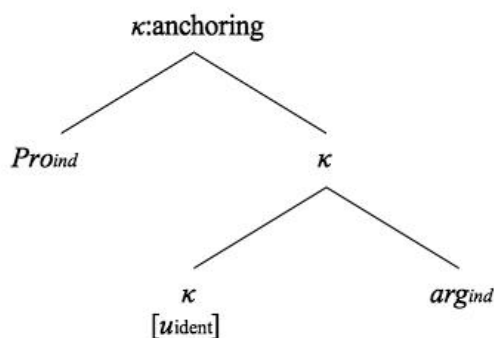
Figure 3: Nominal anchoring (Wiltschko 2014: 209)

In the specifier we have a variable, *Pro_{ind}*, which establishes contextuality. The valuation of [$u$ident] can be realised with different valuation processes (as presented above, i.e. m-, pred- and f-valuation).

Pronouns and demonstratives are valued in morphological valuation, m-valuation, as their anchoring function is (somewhat) independent from their immediate syntactic context. Case is an instance of (syntax-)dependent anchoring, and either valued by pred- or f-valuation. Which valuation process takes place differs according to the case itself. Wiltschko (2014) regards nominative and accusative as instances of pred-valuation. Case that involves prepositions may be analysed in terms of f-valuation: a functional category above D is responsible for case assignment, that is K(ase) (referring to Lamontagne & Travis 1987).

Nominative and accusative are structural cases, externally valued by an (external) higher head. The different external and higher heads valuing [$u$ident] result in different cases: nominative is the result of valuation by INFL/Tense, while accusative is the result of Aspect (Asp) (cf. Wiltschko 2014: 231). In Wiltschko's account, the differences between the two cases are given by the differences in association, *where* to associate with the spine. Nevertheless, we can observe syncretism between nominative and accusative nominals. According to Wiltschko, this is one form of multifunctionality in the context of case. If we have linguistic items that are fully specified for their value [±ident], the source of valuation (pred- or f-valuation), and the association with the spine, no syncretism appears. Wiltschko (2014) explains this syncretism with underspecification of the *source* of valuation: the valued feature can be spelled out, but due to missing information regarding valuation – either INFL or Asp – the item can be interpreted in either way, nominative or accusative.

To sum up: the Universal Spine, structured in different functional domains, interacts with language-specific linguistic elements resulting in language-specific categories. How, where and when these linguistic elements associate with the spine determine the realisation and interpretation of the element and its categorial status. The spine and its functional domains are category-neutral, i.e. independent of the distinction between 'nominal' and 'verbal'. The universal category *K*:*anchoring* is of special interest for DOM, as Wiltschko locates and subsumes different elements and processes regarding nominals under the anchoring concepts of anaphoricity and deixis; exemplarily she investigates demonstratives, pronouns and case.

The split accounts presented here allow for insights into the internal structure of DPs in a broad sense. Either they distribute functions on different layers of a similar type – multiple D-heads or phrases (Zamparelli 2000) – or they distribute features on different syntactic positions, heads and specifiers and on different lexical and functional elements, e.g. nominals and definite articles (Giusti 2015). Wiltschko (2014) supports Giusti's idea of relating Case to D by assuming a common function, anchoring, that relates case and determination. All these approaches allow for a detailed account on DOM in Corsican, as we will see in Chapter 7.

2.3.2 *'Dative accounts' on DOM*

The traditional classification of case as 'inherent' or 'structural' is based, roughly speaking, on the distinction between case that is licensed by virtue of the configuration (following Chomsky 1988), i.e. structural case, and case that is tied to θ-marking, i.e. inherent case. Inherent case is related to a particular semantic role, or to the lexical properties of the governing head. Structural case is driven by syntactic configurations. It may be defined in different terms: thematic roles, grammatical functions, structural positions, or a combination of the three (cf. Baker 2015: 2). Structural case does not have a one-to-one relationship with either thematic roles or to grammatical function. Within case theory, we also make a difference between morphological case, i.e. overt morphological marking, and abstract case, sometimes also called Case (following Chomsky 1988). This Case is not expressed morphologically, but has theoretical impact in terms of Case Theory; i.e., every nominal argument must be licensed via Case as such. If a nominal argument has no Case, a so-called 'case filter' makes the derivation crash. When languages have morphological–inflectional case this is considered "to be a manifestation of abstract Case" (Haspelmath 2006: 3).

Structural case is assumed to be assigned by agreement, but there are also instances of so-called 'dependent' case assignment. Baker specifies dependent case in terms of a local relation between two NEs, but independent of any agreement relation: "The leading idea is that this type of case is fundamentally a relationship between two nominals within the same local domain, not a relation between an NP and a nearby functional head. Dependent case on this view is thus logically independent of agreement" (Baker 2015: 50).

Beyond the functional core of structural case – to relate nominals to the utterance and/or discourse – we can acknowledge that:

> in some languages case marking interacts with definiteness-indefiniteness or specificity-non-specificity […] in other languages the direct object is marked by accusative (or another marker) only if it is definite (Turkish, Modern Hebrew) […] such phenomena have been taken as indications of a close relationship between case and definiteness. (Alexiadou et al. 2007: 151)

Looking now at dative case; this is "either something in between a structural case and an inherent case, or […] a case that has some structural uses and some inherent uses (depending on language)" (Baker 2015: 12). Wegener (1991), who investigates dative constructions cross-linguistically, also acknowledges the double nature of the dative: certain 'dative phenomena' are somewhat structurally driven, as a result of certain syntactic configurations, while in other contexts dative behaves more

consistently, disregarding syntactic particularities.

Analysis of DOM-marked DOs as dative structures can be argued for in semantic and structural terms. Generally, it is rather difficult to define the semantic and syntactic range of datives (cf. Næss 2012 [2008]: 1). Semantically, dative is associated with the roles of recipients, benefactives, experiencers, goals and purposes. Whether any, or all, roles are associated with one single morphological form, is a matter of cross-linguistic variation. Irrespective of this, the most central roles of recipient, benefactive or experiencer are better fulfilled by nominals with sentient, human referents than inanimates. Næss (2012 [2008]: 4) states, "[t]his core meaning of the dative clearly implies that dative marking most typically occurs on animate rather than inanimate nouns". The relevance of animacy for DOM-marked accusatives has been mentioned already. Due to the semantic roles, dative has often been called a semantic case, in contrast to structurally driven cases such as nominative or accusative. But dative may also be defined in structural terms, "where its use can be related to a structural position, namely that of indirect object of ditransitives" (Næss 2012 [2008]: 5). Indirect objects are consistently (morphologically dative) marked, and most often even overrule DOM-marking in a given (ditransitive) situation.

Against this backdrop, I present the following *syntactic* accounts, which analyse DOM-marked DOs in terms of dative marking.

Torrego (1998) is the first to offer a formal syntactic analysis of DOM where DOM is related to the syntax of dative structures. She analyses transitive and ditransitive constructions in Spanish, with morphologically *a*-marked accusative and dative objects. Originally starting with Spanish DOM-marking patterns, she goes on to include dative objects in her analysis, as certain *a*-marked accusative objects and datives share properties if they appear with a certain class of verbs (cf. Torrego 1998: 8).

With regard to *a*-marked objects, Torrego (1998) distinguishes three different types (see Table 9): two *a*-marked accusatives, or DOs (she does not consider unmarked accusative objects in detail), and one *a*-marked dative that is the indirect object (IO). Torrego reframes the attributes 'structural' and 'inherent', traditionally assigned to accusative and dative case, respectively. Some of the *a*-marked DOs behave like indirect, or dative objects. With respect to other properties, all *a*-marked accusatives behave the same. How the properties are aligned, is represented by their position in the table, i.e. their left- or right-orientation within the middle column:

Table 9: Torrego's (1998) characterisation of three types of *a*-marked objects

a-marked object type 1	*a*-marked object type 2	*a*-marked object type 3
with atelic verbs	with telic verbs	
accusatives	accusatives	datives
DO	DO	IO
structural Case	inherent Case	inherent Case
DP	PP	PP
?[12]	SpecvP	SpecvP
a = dative preposition	*a* = dative preposition	*a* = dative preposition
not affected	affected	affected
accusative clitic	dative clitic	dative clitic
word order	word order	word order
accusative pronouns	accusative pronouns	dative pronouns
passivisation	passivisation	passivisation

According to Torrego, the *a*-marking of accusative objects follows two different case strategies, which are the result of the verb classes the respective verb belongs to (cf. Torrego 1998: 54). With telic verbs the object is affected by the verb action. Consequently, accusative is considered an inherent case and the *a*-marker, being a preposition, is heading a Preposition Phrase (PP). By contrast, with atelic verbs the object is considered less or not affected by the verb action. Here, accusative is interpreted as a structural case and the object is structurally a DP. Thus, Torrego argues that *a*-marked accusatives are either structural or inherent cases: "when it comes to Spanish and similar languages, there are reasons to question the idea that accusative is always structural Case. This holds, in particular, for objects in what I call marked accusatives" (Torrego 1998: 7).

The distinction between inherently and structurally *a*-marked accusative objects also correlates with their degree of affectedness and their doubling strategy, as with dative objects: inherently case marked accusatives are doubled by dative clitics (which in turn function as structural case

---

[12] Torrego (1998) does not specify if there are any DOs that may remain *in situ*, i.e. in CompV, that will be *a*-marked nevertheless.

assigners[13]). Structurally marked accusatives already bear structural case and are not necessarily doubled by a clitic. Also, the raising to Spec*v*P is correlated with this distinction: it concerns only inherently case marked accusative and dative objects. Torrego does not explain where the *a*-marked objects (which are not raised to Spec*v*P) remain, nor where they are located, nor how they get case assigned, in the event that they remain *in situ*, i.e. CompV.

As for the nature and position of the *a*-marker, Torrego (1998) considers it a functional category, which she calls a dative preposition. Its feature bundle varies cross-linguistically (cf. Torrego 1998: 25). Despite its status as a functional element, *a* does not always head a proper projection: *a*-marked accusative objects can either be DPs or PPs. According to Torrego (1998: 279), the varying feature composition of *a* corresponds to categorical differences of *a*. In general terms, the *a*-marker as a functional category is "on a par with T [Tense] and *v*, both with respect to the D feature of T and *v* (Chomsky 1995) and with respect to agreement features" (Torrego 1998: 168). According to Torrego, *a* bears a [D] feature in the same way as T and *v*. Nevertheless, Torrego remains vague about the position of *a* if it does not project a PP.

In terms of syntactic trees, this may result in the following structures (see Figures 4 and 5):

---

[13] Torrego states that objects with inherent Case show similar structural requirements for case checking – when no structural Case marking head is available in the structure, clitic-doubling is required: "[T]he doubling clitic is an inflectional element with Case, and […] it can provide the structure with the structural Case feature that is missing from the verb. In other words, I take doubling clitics to be structural Case feature providers", without claiming, however, "that an inherently Case-marked argument always has to be realized in association with a structural Case-marking head" (Torrego 1998: 8, 9). Thus, clitic-doubling is not obligatory.
Parametric variation of functional categories, such as T and *v*, is the reason why some inherently Case marked arguments enter into a structural Case relation with a functional head. According to Torrego, Spanish lexical datives have a double Case requirement for which *v* is responsible. If there is no *v* available to fulfil the double Case requirement, the dative clitic becomes obligatory, whereby "the dative clitic and its double, the argument, originate in a phrase headed by a phonologically null dative proposition" (Torrego 1998: 9).

a-PP
??

[...]

*v*P

Spec*v*P
PP

[...]

P°[D]
a

DP

D°
[...]

NP
[...]
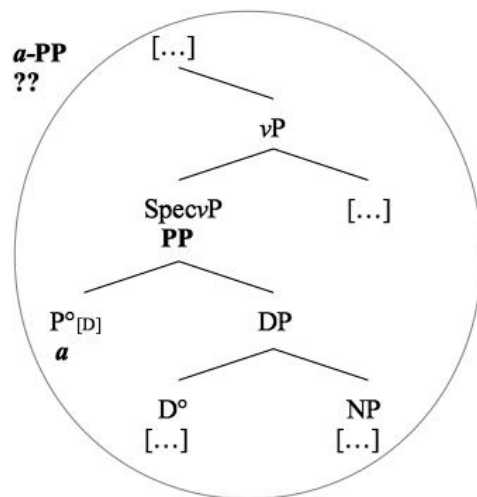
Figure 4: Possible a-marked PP (cf. Torrego 1998)

a-DP₁
??

[...]

[...]

DP

[...]

D°
a

NP
[...]

a-DP₂
??

[...]

[...]

DP

[...]

SpecDP[D]
a

D'

D°
[...]

NP
[...]
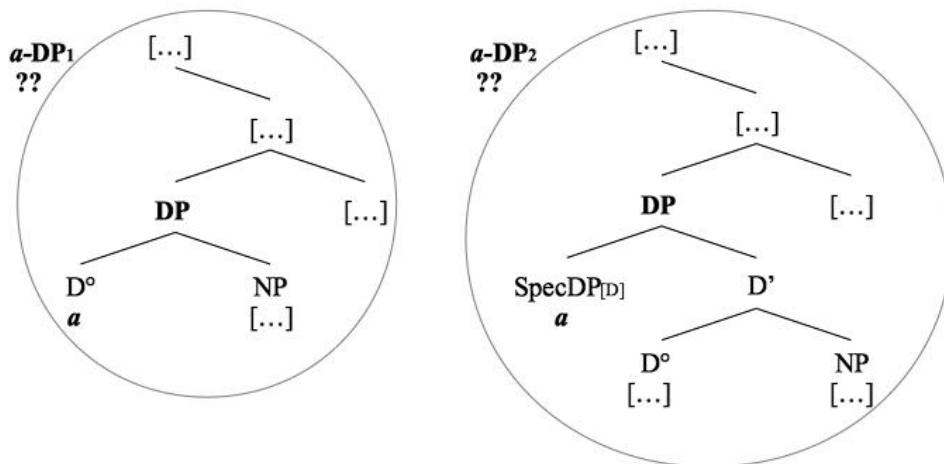
Figure 5: Possible a-marked DPs (cf. Torrego 1998)

The inner structure of different *a*-marked accusative DPs is not further explained by Torrego: the *a*-marker can be located in a D-head or a specDP. Torrego also does not address the relationship between *a* and D-elements. Instead, she links determiners (rather than definite articles) and case, based on morphological distinctions: "Morphological Case characterizes a group of nominals that are superficially disparate. Presumably, the determiner is the part of the nominal responsible for this apparent diversity" (Torrego 1998: 55, 56). In this way, she relates case to the definite article, similar to Giusti (2015).

In summary, Torrego offers a new perspective on accusative case, traditionally regarded as a structural case. To account for the different syntactic behaviour, she distinguishes two types of *a*-marked accusatives: marked either by inherent or by structural case. The resulting implications remain rather sketchy – for example, it is not obvious where the *a*-marker is hosted within in the DP (not PP) structure.

The second dative account is provided by Rodríguez Mondoñedo (2007). He refers explicitly to Torrego's (1998) account, as "the most comprehensive study of Spanish DOM" (Rodríguez Mondoñedo 2007: 156). Like Torrego, Rodríguez Mondoñedo (2007) also assumes a similar structure for *a*-marked accusative objects and dative objects, i.e. a dative phrase. Unlike Torrego, his structure applies to all *a*-marked DOs; he does not distinguish different *a*-marked accusatives. The *a*-marked DOs rise to a dative phrase located above vP where they get case-assigned, i.e. *a*-marked. Rodríguez Mondoñedo (2007) holds the traditional idea that agreement and Case are correlated: a successful matching of φ-features between probe and goal results in Case assignment (following Chomsky 2000; 2001 and *passim*).

His points of departure are existential constructions in Spanish with the verb *haber* and DOM. Rodríguez Mondoñedo (2007) pursues the idea that v of *haber* (v_{haber}) has only one feature, namely [number], but no other φ-features; the feature bundle of v_{haber} is incomplete. As a consequence, v_{haber} cannot license nominals bearing more than [number].

Commonly, little v is considered responsible for accusative case assignment. For DOM-marked DOs, Rodríguez Mondoñedo (2007: 298) assumes a different case assignment process: DOM-marked DOs are not accusative, but dative marked. With this assumption he follows Bossong (1991) and others who assume that both the *a*-marker of DO and the dative marker are the same in Spanish as well as in many other DOM-languages. In terms of movement, that means *a*-marked DOs move into a DatP. The movement is triggered by an unsuccessful agreement process between v and the DO, as v_{haber} is considered a defective head, having only a [number] but no [person] feature. Among the nouns in DO-function, those having a complete φ-feature bundle also bear a [person] feature, and are not able to agree with this defective v. To successfully agree, they must move to another agreement and case assignment position, to an "additional case checker". In Rodríguez Mondoñedo's (2007) account this is a dative head "able to value the case of [a] φ-complete nominal" (Rodríguez Mondoñedo 2007: 4). As a result of a successful agreement, the DO gets *a*-marked within DatP.

Thus, Rodríguez Mondoñedo generalises the idea of an incomplete v with regard to the φ-features: "[S]mall v is φ-incomplete in all Spanish transitive sentences, and therefore that φ-complete

nominals cannot check case against it" (Rodríguez Mondoñedo 2007: 162).[14]

Rodríguez Mondoñedo locates his DatP above *v*P and considers *v* to head a phase (following Phase Theory, cf. Chomsky 2001). To avoid elements with unchecked features being conveyed to spell-out (causing the derivation to crash), these elements must move to the left edge of the phrase. The left edge is not affected by spell-out, but its content is accessible to the next phase, i.e. the next piece of discourse. For DOs with unchecked features, that implies that they must rise to Spec*v*P. From here the nominal may "establish a new Agree relation with a Dative head" (Rodríguez Mondoñedo 2007: 5) and get *a*-marked (see Figure 6)).
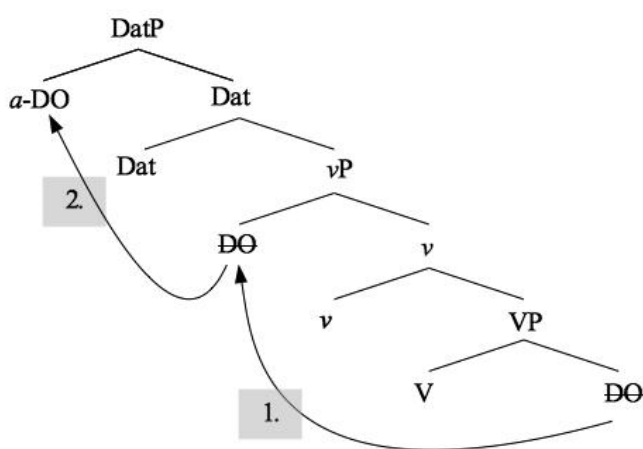
Figure 6: DO-raising (cf. Rodríguez Mondoñedo 2007: 5)

With regard to case marking in more general terms, Rodríguez Mondoñedo, like Torrego (1995), challenges the traditional view of dative being an inherent case and accusative a structural case. Torrego (1995) assumes that there are certain instances of inherent accusatives with regard to *a*-marked DOs. Rodríguez Mondoñedo chooses the other line of argument and assumes that dative case "at least for DOM objects, is structural and not inherent" (Rodríguez Mondoñedo 2007: 186). Hence, he assumes two distinct dative assignment processes: one is inherent, relating to indirect (dative) objects; the other is structural, relating to DOs with *a*-marking, i.e. DOM-DOs. He captures the described peculiarities of dative case (see section 2.3.2, above), being both inherent and structural case.

---

[14] In Spanish, full lexical finite verbs usually inflect for more than just number; besides tense, modus etc., they also inflect for person. Thus, it remains unclear to me, how a finite verb can be inflected for person but not bear [person], while it is inflected for number and bears [number].

Beyond that, Rodríguez Mondoñedo analyses the internal syntax of the DO. He departs from a split KP containing a split NP; see the following structure (10):

(10)   **Rodríguez Mondoñedo's KP**: [$_{KP}$ K [$_{DP}$ D [$_{nP}$ n [$_{NP}$ N]]]]

For single phrases, Rodríguez Mondoñedo stipulates the following properties: K bears an unvalued case feature, K$_{[uCase]}$. For DOs, this unvalued case feature can have two values, one of which is dative; this results in *a*-marking of the nominal, and happens when the DO-nominal enters in a feature checking relation with the dative head. This also applies to IOs. Certain DOs check case with *v* and will remain morphologically unmarked. According to Rodríguez Mondoñedo this concerns inanimate DOs lacking a [person] feature. Animacy is related to [person] via the NP-structure; it motivates the NP split – if NP has an animate (human) referent, then *n* bears [person] (cf. Rodríguez Mondoñedo 2007: 188). By implication, Rodríguez Mondoñedo considers nominals that do not bear [person] to be φ-incomplete. In general, Rodríguez Mondoñedo attributes the [person] feature to the feature bundle of the nominal. Giusti (2015) and Wiltschko (2014) proceed to the contrary: they separate [person] from the nominal and locate it at the left edge of the DP.

Rodríguez Mondoñedo's account displays another peculiarity: the existence of two distinct DPs with different feature compositions, concerning [person] and specificity. The differences relate to two distinct D-heads. For Spanish, Rodríguez Mondoñedo (2007: 112 ff.) assumes two D-heads (D and *D) to account for the different case assignment strategies. These cover different elements that may allow for specific, or non-specific interpretation, (D, or *D, respectively). Beyond definite articles this concerns also certain indefinites and quantifiers.[15]

DPs of the type *DP are inherently non-specific and incompatible with [person]. The selection of *D is triggered by a φ-incomplete *n*P, namely when *n*P does not bear [person]. The *DPs check case with *v*.

D (without an asterisk[16]) is underspecified for any values of specificity. It may either be valued as [specific] or [non-specific]. The projection of D is triggered by *n*Ps with human (animate) referents, which consequently also bear [person]. With regard to case checking, *n*P$_{[person]}$ is considered φ-complete. As *v* is φ-incomplete, no case checking/assignment can take place with *v*, but

---

[15]   On the other hand, he denies D the responsibility for interpretation of specificity. According to him, it is not D – as D allows for specific and non-specific readings – but the environment, that forces specificity, e.g. "[a] prenominal adjective forces the specificity of the nominal" (Rodríguez Mondoñedo 2007: 197). But what is the difference between nominals with a *D-layer pre-syntactically valued as [non-specific], and a D-layer whose [non-specific] value is instantiated in the derivation?

[16]   The notation of D and *D is Rodríguez Mondoñedo's own; the asterisk does not correspond to ungrammaticality.

[DP [*n*P[person]]] has to move further on in the structure; these DPs check case with the dative head.

Figure 7 illustrates the distribution of the features and the processes within Rodríguez Mondoñedo's KP.
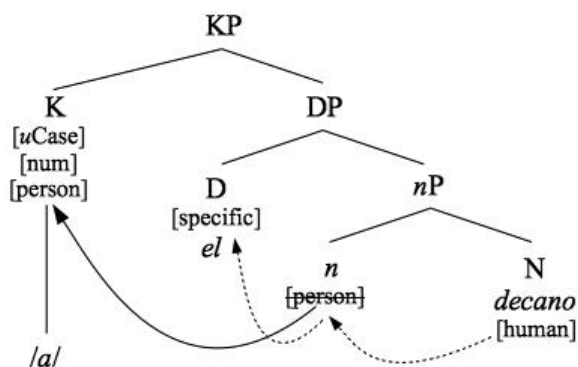
Figure 7: Rodríguez Mondoñedo's (2007:189) KP-structure of an a-marked DO, decano 'dean'

The animacy value [human] of the noun *decano* triggers [person] in *n*P; this in turn triggers the projection of a DP (instead of *DP), which is valued as [specific] due to [person]. By feature percolation (for the mechanism in detail, see Rodríguez Mondoñedo 2007: 189), all feature values available inside KP are 'collected' under K to be accessible for further agreement processes. KP bears [number] ([num])and [person] and is considered to be φ-complete. Therefore, it will not be case checked within *v*P, as *v* only bears [num]. Hence, the KP moves outside *v*P to check case with the DatP, and gets *a*-marked.

Pursuant to the idea that animacy triggers person, one way to 'translate' ontological animacy into morphological and syntactical terms is a [person] feature. The idea can follow two lines of thinking: either we assume that all [human] nominals trigger or bear person, and, by implication, items which bear [person] are unambiguously [human], such as personal pronouns (cf. Rodríguez Mondoñedo 2007: 165); or, in the tradition of discourse participation, we assign [person] to nominals referring to participation – *participants* bear [person], whereas *non-participants* are at most underspecified for [person], and may even not bear it at all. 1st and 2nd person are identified as participants, while non-participants refer to 3rd person. With regard to their properties, participants are [specific] and [human] and uncontroversially bear [person] too. Non-participants show a broader variety: they may be either [specific] or [non-specific], as well as [human], [animate] or [inanimate].

Rodríguez Mondoñedo suggests that "at least in some languages, non-participant nominals that are both specific and animate receive a [person] feature" (Rodríguez Mondoñedo 2007: 34).

According to the 'participation account' and in the terminology of Benveniste (1956), [non-person] is often ascribed wholly to 3rd person nominal items. If [person] is related to DOM, a [person] feature triggered by the animacy value [human] of the nominal in question (irrespective of participation) seems very attractive if these DOM-systems are sensitive to humanness.

The distribution of features within the DP and the resulting processes, as presented above, do not easily account for all *a*-marked DOs in Spanish. The conception of certain nominal items, e.g. indefinites such as *nadie* 'nobody', remains problematic. In Spanish it concerns certain negative quantifiers, indefinites, and interrogatives, namely *nadie* 'nobody', *alguien* 'somebody', and *quien* 'who'.

To account for these *a*-marked DOs, Rodríguez Mondoñedo decomposes the lexical items into single morphemes: *nad-ie, algu-ien* and *qu-ien*. He relates the *-ie(n)*-morpheme to animacy, i.e. [human], and their 'lexical base' to [non-specific]. In the tree-structure shown in Figure 8, the [non-specific] lexical base is hosted in *D.
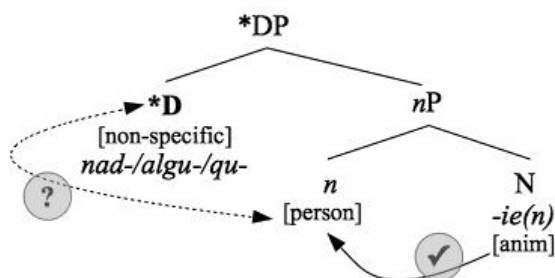
*Figure 8: *DP for nadie, alguien and quien (cf. Rodríguez Mondoñedo 2007)*

Due to the animacy value [human] in N, lower-case *n* bears [person]. Usually, this should result in projection of a D and not a *D-layer. But the inherent value of the above-listed lexical items is [non-specific]. It determines – in Rodríguez Mondoñedo's terms – a *D-projection. The original process seems to be 'overridden' by the direct insertion of the lexical base into *D: here, [person] does not trigger the projection of the D-layer, but of the the locally inserted morpheme. Rodríguez Mondoñedo does not explain why D, a functional head, can host lexical items. While the internal structure of a DP/*DP may be accessible and visible in morphological terms, it is not accessible for syntactic processes: *nadie, alguien* and *quien* are considered "complex lexical item[s]", for which Rodríguez

Mondoñedo assumes an atomic structure: "the hierarchical structure of the features inside of it is not visible from outside" (Rodríguez Mondoñedo 2007: 203). This is represented in Figure 9, as follows:
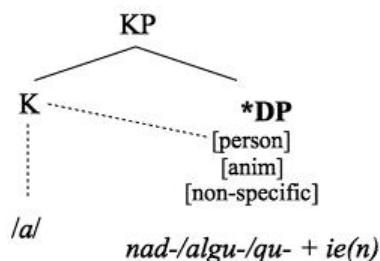
Figure 9: Rodríguez Mondoñedo's KP (2007: 203)

With regard to the nominal structure of DOs, *a*-marked or not, Rodríguez Mondoñedo (2007) invokes two more constructions that are repeatedly discussed in the context of Romance DOM: bare plural nouns and DOs in ditransitive constructions.

With regard to ditransitives, Torrego (1998) explains the difference between possible DOM-omission and obligatory application in ditransitive constructions, by means of different case checking processes related to the different kind of verbs. Rodríguez Mondoñedo (2007: 224) shows that both, i.e. possible omission and obligatory application of DOM-*a* may happen with the same verb: The omission is prevented under certain circumstances, such as complexity of the objects, e.g. DOs constructed from more than one lexical item, and pronounced with special intonation or a change in word order, for example, if the objects are left-dislocated and clitic-doubled.[17]

To summarise, Rodríguez Mondoñedo follows the more traditional view with respect to the distinction between inherent and structural case. He offers two dative cases, one of which is inherently driven, the other being instead a structural case and concerns the *a*-marked DOs. Relating [person] to the animacy value of the noun is very useful for translating ontological animacy into morphosyntactic terms, and to account for DOM-marking strategies triggered by humanness.

The last dative account that I will discuss here is from Ormazabal and Romero (2013). Their account can be grouped with the other dative accounts, but in fact offers a completely different approach; they deny unmarked DOs any case at all. Thus, on the one hand, Ormazabal and Romero (2013) stand in the tradition of the 'DOM-Dative' accounts, where the DO and the IO compete for the same

---

[17] Rodríguez Mondoñedo (2007: 225–256) offers a details analysis of DOM in the context of ditransitives, related to linearisation processes and rules.

syntactic position (cf. Torrego 1998; Rodríguez Mondoñedo 2007). On the other hand, they call for *caselessness* to account for the distinction between 'marked = Case checked', and 'unmarked = Caseless nominals'. Caseless nominals in the context of DOM have been named already by Danon (2006) to analyse Hebrew DOM. For Ormazabal and Romero, Case and agreement concern different checking relations. Noticeably, the interpretative effects concerning animacy, referentiality or topicality are secondary to their approach, and result from structural conditions: "[T]he set of objects that are subject to DOM is semantically heterogeneous enough not to attempt an approach based strictly on semantic primitives" (Ormazabal & Romero 2013: 222).

In structural terms, certain DOs and IOs compete for the same syntactic position. They assume a single *a*-marker, at least for a certain subgroup of *a*-marked DOs and IOs: those which are clitic-doubled. See, for example, the following sentence (5):

(5)   ***Le** echaron agua **al** vino*

     'They poured water in the wine.'   Ormazabal & Romero (2013: 223, 4b)

Here, they state that the accusative/dative distinction in Spanish does not correspond to two different structural relations in the functional layer. Rather, the morphological distinctions traditionally associated with dative and accusative in the pronominal system of Spanish, distinguish between objects that have entered into an agreement relation and elements that have not (Ormazabal & Romero 2013: 223).

Nominally different – but very similar in its basic conception to Torrego's unified treatment of (a certain class of) *a*-marked DOs and IOs, and also to Rodríguez Mondoñedo's unified class of *a*-marked objects – is Ormazabal and Romero's (2013) consideration of the distinction between accusative and dative in the context of DOM as irrelevant, as the *a*-marking concerns the same structural position. For them, "the real division is then between objects that enter into a case checking relation in the specifier of *v*P and those that do not" (Ormazabal & Romero 2013: 233).

A-marked (direct and indirect) objects enter into an agreement and case checking relation, while unmarked DOs neither agree nor get Case assigned. Although Ormazabl and Romero ascribe the negation of *both* processes to unmarked DOs, they understand the relationship between Case and agreement as somewhat heterogeneous, contrary to what is commonly assumed:

[L]anguage-internal and crosslinguistic empirical evidence strengthens the idea that only
a subset of objects is formally Case-licensed, and any detailed analysis concerning the

empirical scope of the Case-Filter, on any of its versions, yield a rather non-uniform and, at least partially incoherent set of objects. (Ormazabal & Romero 2013: 233)

Case checking of *a*-marked objects takes place in Spec*v*P, the traditional case checking position since introduction of little *v*. Here, Case of both DOs and IOs is checked. For ditransitive constructions this results in the previously-discussed competition between DOs and IOs for the same syntactic position, and thus, for the *a*-marking. But two *a*-marked objects cannot occure within the same clause.

Empirical data show that the distinction is not so simple. There are ditransitives, where the IO is obligatorily marked, while the DO may or may not be marked, i.e. two *a*-marked objects may appear; Whereas in sentences where the IO is doubled by a clitic, the DO cannot be *a*-marked: **Les enviaron a Mateo a los doctores* 'They sent the doctors Mateo' (cf. Ormazabal & Romero 2013: 224, 236). For these authors, it is the doubling clitic on the dative that seems to prevent *a*-marking on the accusative, rather than restrictions on the object in a proper sense. This may be related to Torrego's idea of clitics being related to case checking and assignment (see above).

Like Torrego (1998), Ormazabal and Romero (2013) also remain vague about the internal syntactic structure of the DO. Basically, they conceive *a*-marked DOs as DPs, the same as *a*-marked IOs. On the other hand, they mention a PP construction for certain datives in ditransitive constructions (cf. Ormazabal & Romero 2013: 223). The *a*-marker itself is not properly classified.

To summarise, in all 'dative accounts' of DOM, certain accusative objects compete with dative objects in structural terms. All accounts refer to ditransitive constructions where only the dative consistently bears *a*, while the *a*-marker on the accusative object is often omitted even if properties of the nominal or DP would trigger *a*-marking in a transitive construction. For Torrego (1998) and Rodríguez Mondoñedo (2007), case marking takes place at a position above Spec*v*P, the structural position where case checking traditionally takes place. For Ormazabal and Romero (2013), Spec*v*P itself is where case checking takes place. A second observation concerns case checking and assignment processes in general. For Torrego (1998) and Rodríguez Mondoñedo (2007), all DOs require case checking and assignment, i.e. they do not question the case requirement. But they do question the traditional distinction between structural and inherent case, and accusative and dative case. By contrast, Ormazabal and Romero override the interdependence between checking and assignment, assume them to be independent processes, and allow for nominals to remain caseless (and unmarked), if case checking and case assignment do not occur.

2.3.3 *'Accusative accounts' on DOM*

The following accounts are characterised by the basic distinction between accusative and dative objects, accusative case being a structural case. In terms of the processes by which agreement is assigned, all accusative accounts on DOM follow traditional assumptions about case assignment, i.e. every nominal requires case.

In the first of these accounts, Brugè and Brugger (1994: 11) categorise the *a*-marker as a functional head (F) selecting for a DP. Its case value will be assigned by V. The *a*-marker works as a dummy preposition: it is the realisation of accusative case assigned by the verb V to its internal argument. "[W]e propose that *a* occupies the head of a Functional Projection, FP, conceived of as a Case Projection, which can select a DP" (Brugè & Brugger 1994: 11); see Figure 10 below.
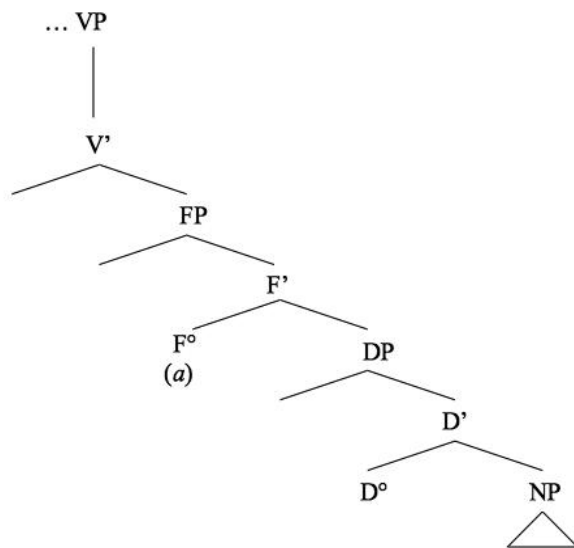
Figure 10: Brugè & Brugger's (1994: 11) syntactic structure of DOM

Furthermore, Brugè and Brugger (1994: 11) assume that FP is always projected in syntax; it is the locus of case assignment by the verb to its internal argument. In comparison to upcoming *v*P-shell analyses, Brugè and Brugger's account already represents one of the main strands of syntactic DOM analyses: *a* heading an own functional category of its own, of the type FP, later also CaseP, or KP. As we will see later, López (2012) also operates with a functional projection above DP, called KP.

Brugè and Brugger ascribe *a* to its own functional projection, but they conceive it as "an extension of the Noun itself, and that, for this reason the head F contains some nominal features, such as [± Accusative] and [± Animate], underspecified for its positive or negative values (cf.

Grimshaw 1990)" (Brugè & Brugger 1994: 11).[18] Subsequently, *a*, which is specified for [acc] and [anim]/[human] will be realised in the head of the functional projection, F°, if the nominal is equally specified for these values.

The distinction between a morphologically marked and an unmarked DO, corresponds to accusative and partitive case, so Brugè and Brugger: "We assume that Spanisch Direct Objects can be marked either with Accusative Case, or with Partitive Case" (Brugè & Brugger 1994: 35). In this sense, unmarked bare nouns are marked with partitive case. This is different from other accounts on DOM that we have seen so far; unmarked DOs are considered accusatives in structural terms, independent from overt marking, with the exception of Ormazabal and Romero's 2013 account, wherein unmarked DOs are considered caseless (see section 2.3.2).

In summary, Brugè and Brugger's account is based on some kind of feature checking and/or agreement, regarding animacy values and case, respectively. The DOM-marker *a* is specified for [acc] and [human] and will be realised in F°, if the nominal in question is equally specified.

While Brugè and Brugger (1994) assume two cases for DOs, i.e. accusative and (morphologically zero marked) partitive, López (2012) postulates only accusative case for DOs, but two different underlying processes to satisfy the case requirement:

> In the present analysis there is only one type of accusative Case, and there are two alternative pathways to satisfy it, each coming into play depending on the composition of the nominal phrase […] In my approach, there is one type of Case which is assigned by incorporation unless the structure of the nominal phrase makes it impossible, in which case XP-movement takes place. (López 2012: 51)

Indeed, López (2012) assumes two different constructions that may allow for accusative case assignment, but his division does not correlate with *a*-marking or non-marking of the DO. Briefly, *a*-marked DOs must project a KP and rise to *α*P (shown as P1 to P2, in Figure 11). Unmarked DOs may either remain *in situ*, i.e. CompV (P1), where they incorporate into the V and the incorporation structure moves along to satisfy the unvalued Case feature, or may equally project a KP and rise to *α*P (P2), but their K remains phonologically empty.

---

[18]   Similarly, Giusti (2015) also perceives case as part of the nominal; see section 2.3.1.

Before presenting López's different DO-structures, I will introduce his syntactic configuration, with each of the different positions – P1, P2, and P3 – for the DO. See the following tree in Figure 11:
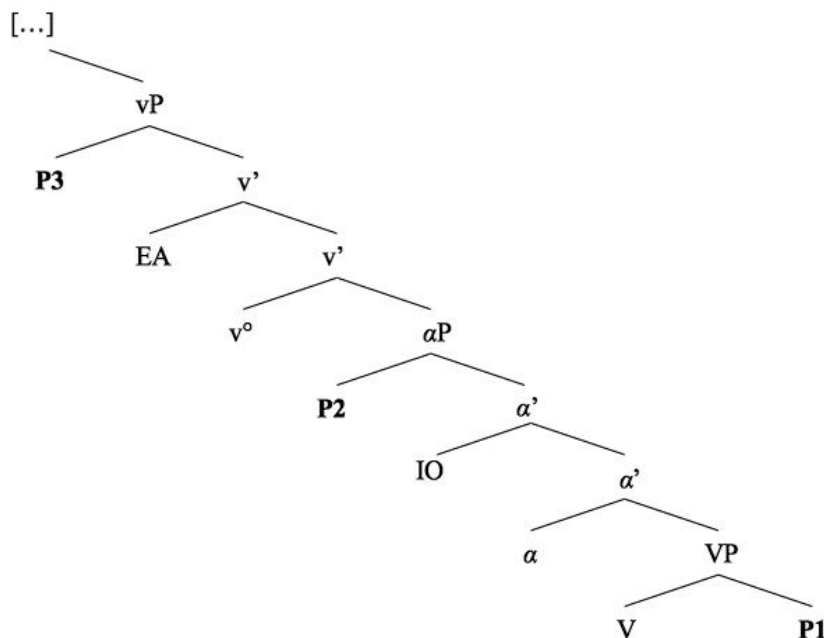
Figure 11: López' (2012: 45) positions of DOs in the clause structure: P1, P2, P3

In P1 all DOs are base-generated and this is the locus of unmarked DOs that remain *in situ*. These DOs are [non-specific] and take narrow scope. Potentially, *a*-marked DOs move on to P2 where they may, but do not have to be, *a*-marked. These DOs can be [specific], but are not necessarily so. They may take either wide or narrow scope. By contrast, P3 instead is not primarily related to Case assignment: "Constituents that occupy P3 and establish a dependency with *v* are subject to an interpretative rule that turns them into discourse anaphors" (López 2012: 45). Contrary to Giusti (2015) and Wiltschko (2014), López does not formulate the relationship between case and discourse.

If we compare the structure in Figure 11 with those presented earlier (cf. Brugè & Brugger 1994; Torrego 1998; Rodríguez Mondoñedo 2007), we can observe that López's potentially *a*-marked DOs move to a position lower than the external argument, the specifier of an intermediate projection *α*; while Torrego's (1998) and Rodríguez Mondoñedo's (2007) DOM-DOs are situated in (the only) Spec*v*P – the position which López reserves for discourse anaphors.

DOs which project into DPs, Number Phrases (#Ps), or NPs, are taken to be incorporated into V (see option 1 in Figure 12). The whole VP will move on for feature valuation and Case assignment (see López 2012: 47, 48 for the procedure). DOs that can be *a*-marked project into a KP (option 2 in Figure 12). The projected K-layer prevents incorporation. KPs move to Spec*α*P as their unvalued Case feature [*u*C] can not be valued *in situ*. This is illustrated in Figure 12).

In Spec*α*P, [*u*C] is valued and K is realised as /*a*/ if the NP and/or its referent bear the required feature
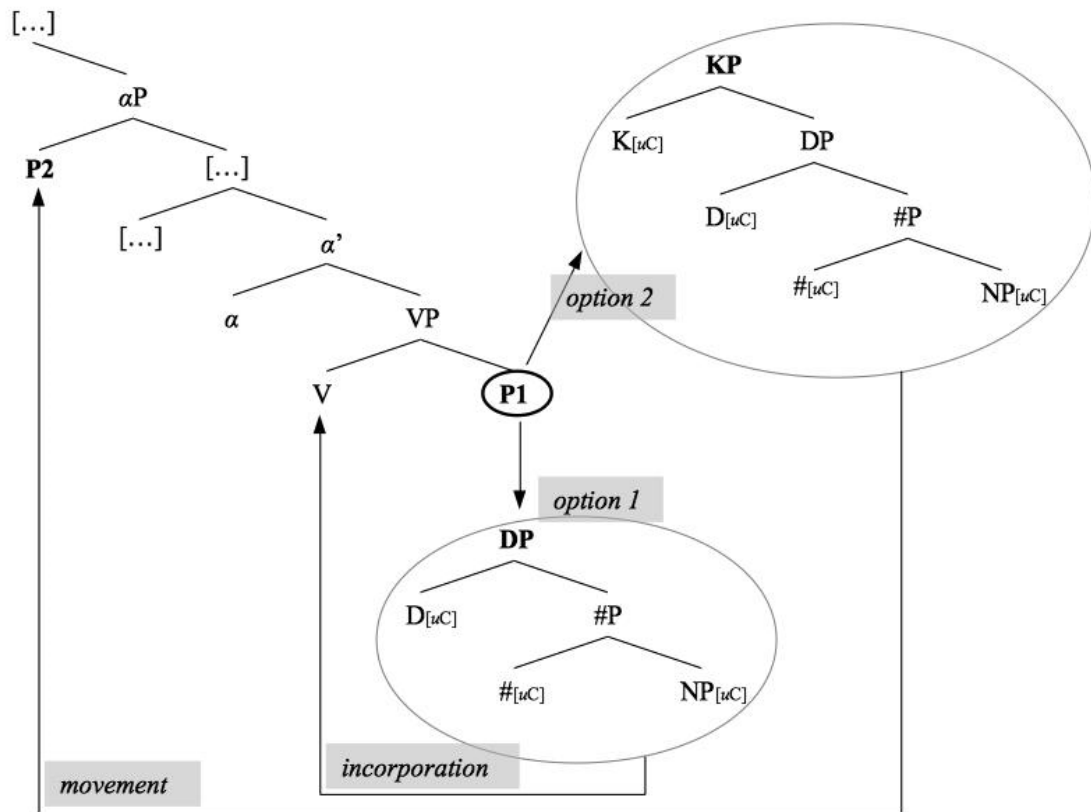
Figure 12: DOs' projections and their structural behaviour (cf. López 2012: 47, 48)

values, e.g. [human] and/or [specific]. If the feature bundle of K and NP in its complement position do not match, K has no spell-out. This is the case for many inanimate nominals in Spanish. Language-specific properties may further constrain the appearance of DOM:

> These properties are in effect features that can be found in the immediate local environment of K. Thus, the spell-out of K can be influenced by (i) features of the nominal phrase selected by K, as well as by (ii) features of α, (iii) features of v, and even (iv) features of the EA [external argument] selected by v. (López 2012: 60).

With regard to the features of the DP selected by K, the spell-out rules of K are determined either by

properties of D (i.e. referentiality, definiteness, specificity) or of the NP within the DP (i.e. inherent features concerning the referent of the nominal, e.g. animacy). Nominal classes also play an important role. In some languages, only a certain nominal class get *a*-marked, as is the case for Catalan personal pronouns; here, K is spelled-out only if the DO is a personal pronoun (see López 2012: 61–64 for detailed insights from different DOM systems cross-linguistically).

About the nature of K, López (2012) states the following:

> Let us assume that K is morphologically specified as a prefix of D. Further any copy of K will have this property. Thus, K cannot incorporate into V because the copy of K in V could not prefix to D. Moreover, the Head Movement Constraint (Travis 1984) ensures that D cannot incorporate into V across K. Since the path of incorporation is closed, KP cannot receive Case from v or any other higher Case assigner, given the strict locality on syntactic dependencies […] KP must move to the closest available specifier, Spec,α. In that position, the marked object can be probed by v and receive [accusative]. (López 2012: 49)

This explanation leaves open many questions. Where does the functional category K come from? It might not be related to Case requirements, as all nominals inserted in CompV require Case; according to López, the incorporated DOs also move along due to case requirements. It cannot be definiteness or referentiality either, as these are already related to the D-projection. What triggers the projection of a KP at all? Is K related to inherent features of the DO or DO-referent? This would makes animacy a good candidate: if the noun has already projected a DP due to individuation, referentiality, definiteness, or the like, animacy comes into play and allows for a KP. Hence, structurally, animacy would be postposed with regard to definiteness. This hypothesis meets the observation that all accounts working with a Case-related functional category (e.g. Giusti 1993, 1995a; Brugè & Brugger 1994; López 2012) assume a DP below KP/CaseP/FP. On the other hand, bare plurals referring to humans have a mass denotation. While nominals with individualised human referents may behave differently from other nominals with regard to morphology, syntax, or morphosyntax, [human] bare plurals are identical or at least similar to other mass-denoting nominals, almost as if humanness would lose its morphosyntactic impact in certain plural contexts. This also meets Yamamoto's (1999: 4) observation that plurality sometimes weakens the role of animacy because the identity (identities) of

the referent(s) can be blurred.[19]

But we have also inanimate DOs bearing *a*, cases that can not even be explained with inferred animacy (see section 2.1). Thus, the *projection* of KP seems to be independent of any animacy feature, whereas the *spell-out* of its head, i.e. K, may be but is not necessarily triggered by animacy.

K may be motivated by specificity: "The role of K is to lift the DP to an *<e>* type […] Thus, K is necessary for specific readings" (Fábregas 2013: 52). Brugè and Brugger (1994) related the distinction between *<e,t>* and *<e>* type to two distinct cases, partitive versus accusative. For Fábregas (2013) type-shifting is related to specificity. Zamparelli's (2000) determiner system, operating with two subordinated DP-layers, follows a similar approach; not all DPs are of type *<e>* (see section 2.3.1). Fábregas' argumentation for K, wherein its function is to shift the object from *<e,t>* to *<e>*, counters with the common assumption that D is already responsible for changing the nominal from *<e,t>* to *<e>*. D° elements are type-shifters, as stated by Alexiadou et al. (2007):

> [P]redicates are of the semantic type <e,t>, individuals are of the type <e>. Articles (and more generally determiners) are of the semantic type <<e,t>,e>: they are 'functions' that take a predicate, which is of the semantic type <e,t>, and yield an individual, which is of type <e>. (Alexiadou et al. 2007: 64)

If KP is related to specificity, then specificity is tied to a syntactic position or configuration, which corresponds to Diesing's (1992) account.

> DOM is often related to specificity: in many languages, indefinite objects with such a marker are (or can be) specific, while unmarked objects are not. Spanish is one such language […] With accusative *a*, it [the indefinite direct object] can have a specific reading. Without accusative *a*, it can only be non-specific. (López 2012: 10)

While this argumentation is reasonable for indefinites in relationship to specificity, it raises questions in the context of definite DOs. Do they require *a* to be interpreted as [specific]? It is obvious that there is an interrelation between specificity, animacy and *a*-marking in Spanish, but their interaction

---

[19] This observation is also not surprising from a anthropocentric point of view: humans are considered individuals in the first place. Singularity, referentiality, specificity, definiteness, etc., are features attributed to [human] nouns which are standing to reason.
Ormazabal & Romero (2013: 224) mention – extremely restricted – contexts where nominals with human referents do not allow for *a*-marking, even if they bear a determiner. They relate this to the phenomenon of "de-animation of the NP: [such sentences] are only grammatical with nouns referring to sick people, soldiers, slaves, kids, etc. most felicitous, collectively [*sic*]; that is, animate individuals that may be contextually subject to such a de-animation process".

does not follow a simple structural pattern.

To the general questions concerning K and its projection, we may add more specific questions regarding KP in combination with certain lexical items. As seen earlier (cf. Rodríguez Mondoñedo 2007), negative and indefinite quantifiers often challenge the proposed structures. Likewise it does for López's (2012) account. He opposes Spanish *nadie* 'nobody' and *alguien* 'someone', which refer exclusively to humans and require *a*-marking, to *ningún-* 'no-' and *algún-* 'some-'. The latter two may be *a*-marked in cases of human referents, and unmarked when referring to inanimates.

According to López (2012) this restriction in marking with reference to the lexical item has to do with strict vocabulary insertion rules, as it affects only these two lexical items. "According to many linguists who have written about *nadie* and *alguien*, the obligatoriness of accusative *a* with these words is a consequence of their animacy feature, which somehow trumps their lack of specificity" (López 2012: 63).

López 'translates' animacy into a morphosyntactic feature of certain Ds:

As for regular nominal phrases the noun is [± animate] and the D is neutral with respect to animacy (i.e., the determiners are the same for animate and inanimate nouns), *alguien* and *nadie* arguably include a [+animate] D, inherently or by incorporation of an animate noun (as noted by Rodríguez Mondoñedo 2007). This feature makes *alguien* and *nadie* different from *ningún-* and *algún-*, which are not inherently animate. (López 2012: 63, 64)

For this argumentation, López offers the structure shown on the left-hand side of Figure 13. On the right-hand side I have inserted the aforementioned lexical items by way of demonstration. Under option 1 are *nadie* and *alguien*, considered inherently animate and endowed with an [anim] D. In option 2 are the other two lexical items, *ningún-* and *algún-*, not inherently specified for animacy, and hence adjoined to a 'neutral' D (López 2012: 63, 64) with respect to animacy.

Some questions remain open:
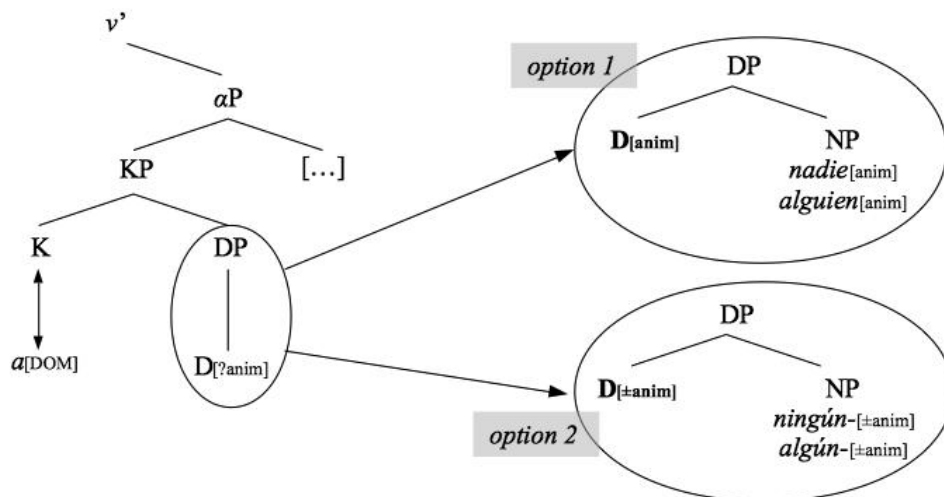
Figure 13:   Left: López' (2012: 64) D[anim] structure
Right: with vocabulary insertion

First, how do these lexical items, categorised as negative and indefinite quantifiers, project a DP, as they are not referential, definite, specific, unique, nor any other property traditionally related to D? How do they overcome these requirements?

With regard to D: how is the D in option 1 valued/specified as [anim]? And, as counterpart to D[anim], how is this neutrality of regular Ds represented pre-syntactically? If we assume processes such as feature sharing/inheritance (cf., for example, Grimshaw 1990; Giusti 2009), D would get the animacy feature of the contained nominal in either way, just the same, as soon as a DP is projected. D°-elements, i.e. definite articles, do not inherently bear an animacy value (see section 2.1, where animacy and its ontological nature are presented in detail).[20]

---

[20]   The Catalan determiner system distinguishes between nominals referring to humans and non-humans by selecting different articles. Hence, animacy may reflect on the D-element simply by selection of its respective definite article. This is an exception among the Romance languages.

Furthermore, is there any phonological realisation of D in options 1 and 2? Or does the NP move to D? Or does D remain phonologically empty, and nevertheless is able to trigger DOM? To avoid a D-layer for these items, a possible option is to allow for 'smaller phrases' in terms of #P, NP or φP (all below DP) to project into KP. According to Fábregas (2013), K can fulfil the essential function of type shifting, to shift the noun from $<e,t>$ to $<e>$ type, a capacity traditionally ascribed to Ds.

Finally, considering the lexical items *nadie* and *alguien*: how and where do they get their animacy value instantiated? López (2012) did not work out in detail the different nature of *nadie, alguien* compared to *ningún-, algún-*, nor the nature of their feature bundle, nor their nominal classification. For the feature bundle, I assume that *nadie* and *alguien* inherently bear [human], as they are exclusively used to refer to humans. For *ningún-* and *algún-* the animacy feature is not inherently specified, but the elements get valued in the derivation by information given in the discourse – if used anaphorically/pronominally, by some long-distance agreement process or some kind of D-linking, and in prenominal position by local agreement with the noun phrase to which they refer. This distinction between 'pronominal' and 'prenominal' use is related to the nominal classifications of these elements: *nadie* and *alguien* can only be used as pronominals, whereas we find *ningún-* and *algún* in different syntactic positions: they can be realised as pronominals whereupon they can be interpreted as synonymous to the first ones, *nadie* and *alguien*, in the event that they refer to humans – *nadie = ninguna/o* and *alguien = alguna/o* – with a difference in gender marking; or they can be used in prenominal positions, *Hoy en clase, no he visto a ninguna chica, solo chicos* 'Today in school, I saw DOM no girls, only boys' and *Yo si, yo he visto a alguna chica* 'I did, I saw DOM some girls'. Somehow this difference has to be encoded; *nadie* and *alguien* may bear an (inherent) feature that prevents their prenominal use, and for which *ningún-* and *algún-* are underspecified.

An explanation may be provided by von Heusinger and Kaiser's (2003) 'anchoring account'. Here, certain items are anchored to the discourse by some element given in the context, to which the nominal in question is related (cf. von Heusinger & Kaiser 2003: 48). How to translate this in structural terms? Either: *nadie, alguièn, ninguna/o*[pron]*, alguna/o*[pron], etc. can project into a KP without a DP-layer, as they do not bear features traditionally related to definiteness and D°. Or: by virtue of the proposed 'anchor-specificity', they project into DP due to their linking to the discourse. In addition, animacy could have some triggering power to project the DP. Structurally, they then conform to all other KPs in having a proper DP.

In summary, the syntactic accounts presented all share a basic tendency: *a*-marking, or in broader terms Case assignment, is related to a proper functional layer. It can be located at different points in the derivation: lower than *v*P, in *v*P or above *v*P. It may be a multifunctional layer, responsible for accusative and dative case under certain conditions, or a mono-functional layer which is specified for accusative case. With regard to the mono-functional layer, it may realised in two ways: either by the overt *a*-marking of the DO, or with no phonological realisation.

The syntactic accounts share the requirement for a DP. The functional projections, whether called KP, FP, *α*P or DatP, can only be projected if there is a DP-layer. Indeed, a missing DP is argued to be the reason for non-marking of nominals, even if they are able to satisfy certain semantic requirements for *a*-marking.

All dative and accusative accounts are concerned with case assignment and where, when and how it takes place. They are much less focused on the DP-structure and the realisation of case on the nominal itself. To account for DOM in Corsican I do not question general case assignment processes and rules on the clausal level. I am interested how the DP and the nominal, i.e. structure and item, interact and how they influence each other regarding the phenomenon of DOM.

3 Corsican

The following chapter offers an overview about the Corsican language and its documentation in the broadest sense. The first half of the chapter is dedicated to issues regarding linguistic differentiation and classification; the second half to the historical background of the island, roughly separated into three periods: he pre-Tuscan, the Tuscan, and the French periods, (up until 1120 AD, from 1120 to the 1768, and from 1768 until today, respectively).

I will start with the diatopical differentiation of Corsican (section 3.1.1), and its classification (section 3.1.2). Almost as old as the study of Corsican itself, the issue of 'original' Corsican has never lost its significance; the relevance of the topic with regard to language data or language elaboration is treated in section 3.1.3. Section 3.1.4 presents (written) language documentation, from the first known documents until today. Empirical studies on Corsican, in particular based on oral data, as found for example in linguistic atlases, are offered in section 3.1.5.

3.1 Differentiation and Classification

3.1.1 *Diatopic differentiation*

The diatopic differentiation of Corsica's language varieties has been one of the principle areas of study since Corsican became the focus of scientific linguistic investigation, in the second half of the 19th century.[21] Emphasis was placed on dialectal studies (among others by Antoine Mattei 1864; Francesco Domenico Falcucci 1875, 1915; Grazidio Isaia Ascoli 1882; Pier Enea Guarnerio 1896; Carlo Salvioni 1916; Clemente Merlo 1925), based on comparative accounts in the fields of phonetics, phonology and lexis.[22] Corsican was then considered a Tuscan(ised) dialect, but a common former Sardinian–Corsican language unity is also observed (cf. Bottiglioni 1957a [1936]; Loi Corvetto & Nesi 1993: 214).

Any assignment of Corsican as 'one language' to one or another language group is impossible because of Corsican's pronounced diatopical differentiation. In general, the variety of the north,

---

[21] The inclusion of Corsican as a separate entry in the *Lexicon der Romanischen Linguistik* (LRL) (Holtus, Metzeltin & Schmitt 1988) is considered a milestone in the history of Corsican linguistics. According to Thiers (1993) it marked the key moment when Corsican was recognised as an independent and official topic of modern linguistics by linguists. It also reflects the recognition in society and policy it gained in the same time.

[22] Even into the 1980s, most empirical studies on Corsican almost exclusively considered its phonetic and phonological characteristics of single lexems. Any other approaches related to this field of study, e.g. intonation or prosody, had been ignored ("Le niveau phonologique du corse, par exemple, a été relativement mieux décrit: mais on ne sait rien, ou pas grand chose, des structures prosodiques, intonatives dont chacun mesure pourtant l'importance, notamment put la reconnaissance ou l'auto-conscience communautaire" [For example, the phonology of Corscian was relatively well described. But little ot nothing was known about the prosodic and intonational structures, the importance of which can be measured by everyone, particularly of recognition or community self-awareness; KAN] (Chiorboli 1987: 99).

corsu supranu or *corsu cismuntanu,* is considered more 'Tuscanised', and stands in closer relation to standard Italian, while the southern variety, *corsu suttanu* or *corsu otramontanu*, is similar to Gallurese, a northern Sardinian variety; see Map 1.

The most current and extensive studies on the diatopic variation of Corsican have been conducted by Dalbera-Stefanaggi (among others 1991a, 1995, 2001); these studies are related to her work on the *Nouvel Atlas Linguistique et Ethnographique de la Corse* (NALC). She takes a focus on phonetic and phonological features – in a diachronic and synchronic view – to reveal different diatopic varieties. Her studies, in particular on vowel accentuation (cf. Dalbera-Stefanaggi 1991a; 2002) reveal four major diatopic areas; see Map 2.

Map 1: Diatopic variation, binary division (Melillo 1977:22)

Map 2: Diatopic variation, quadripartite (Dalbera-Stefanaggi 1991a:493)

In general, Dalbera-Stefanaggi argues for a dialectal continuum, as the subdivided diatopical areas do not display clear borders (cf. Dalbera-Stefanaggi 2002: 6), as opposed to the borders described by Contini for Gallurese (cf. Contini 1987). Thiers (1993: 255) also argues in favour of a dialectal continuum, not only because of the difficulties in drawing the line between single varieties, but also because of the inter-comprehensibility between all Corsican speakers irrespective of their provenience. The classification of the diatopic varieties within a dialectal continuum matches Jean-Baptiste Marcellesi's (1984) and Thiers's (1993) approaches of Corsican as a 'polynomic' or 'polyphonic' language (for more about this, see below).

Corsican is also perceived by Mathée Giacomo-Marcellesi as being subdivided into more than two varieties; her 'variétés micro-régionales' (Giacomo-Marcellesi 1978a: 127) are based on the ecclesiastical-administrative structure of the *pievi*[23] (Fusina 1999: 20). We find lexical differences,

---

[23] *Pievi* is an ecclesiastical administrative unit similar to a parish.

between for example, *kakàru* (Cape Corse and Bastia) and *misiau* (Aiacciu and the south-west) 'grandfather' (Giacomo-Marcellesi 1978a: 237, Fig. 47). And we find differences in the phonetic/phonological domain: for example [u bappu] (in Sartène, Tavaro and Petreto-Bicchisano) versus [u babbu] (Sotta, Bastelicaccia) 'the father' (Giacomo-Marcellesi 1978a: 224).

More recently, morphological and syntactic characteristics have also been taken into account from a diatopical perspective. Marcellesi (1983) describes seven '*indicateurs d'identité*' (Marcellesi 1983: 399): four phonetic characteristics (i.e. nominal ending in -*u*, the palatal affricates /tj/ and /dj/, word-initial consonant mutation, and vocal gradation according to intonation); two morphological particularities (i.e. the definite article system, and the future tense built on *andà* 'to go' + infinitive); and one syntactic phenomenon, DOM.

Chiorboli (1987) considers, among his twenty-two selected morphological and syntactic characteristics: the article system; nouns (including DOM); adjectives and adverbs; possessive constructions; the clitic order of accusative and dative; and relative and interrogative clauses (the features under investigation are listed in tabular form in Chiorboli 1987: 83). Nesi (2002) describes very briefly just six particularities: definite articles, the subject pronoun, clitic order, the impersonal pronoun *omu* 'one', possessive constructions, and DOM (cf. Nesi 2002: 962, 963). Dalbera-Stefanaggi (2002) refers exclusively to the system of nominal endings and the definite article.

Surveys that take syntactic or morphosyntactic structures and features into account describe Corsican as a syntactically homogeneous variety (cf. Bottiglioni 1957c: 121; Casta 2003: Préface; Nesi 2002: 962).

3.1.2 *Language classification*

The classification of Corsican or the Corsican varieties has always been related to the historical and the historic-linguistic background of the island: its classification among the Tuscan varieties alludes to the historical background during Pisa's and Genoa's rule over Corsica. Seen historically, the classification of Corsican as a Tuscan Italian dialect has a long tradition. In line with this, Diez (1836) collocated Corsican along with the Tuscan and Roman dialects among the central Italian varieties. Ascoli (1882: 110, 111), in his *L'Italia Dialettale*, classified Corsican under the "dialetti che si scostano, più o meno, dal tipo schiettamente italiano o toscano, ma pur possono entrare a formar col toscano uno speciale sistema di dialetti neo-latini". Carlo Salvioni likewise alludes to the common linguistic history of Corsica and Italy/Tuscany: "Tutto, in Corsica, lingua, costumi, arte, consuetudini e organizzazioni giuridiche, tutto attesta comunanza antica coll'Italia e più precisamente colla Toscana" (Salvioni 1916: 713) [Everything in Corsica, language, apparel, art, customs, and judiciary,

confirm an ancient community with Italy, or more precisely with Tuscany; translation KAN). Pier Enea Guarnerio also argues in favour of a Corsican–Tuscan history (Guarnerio 1896: 386).

Allocation to the central and southern Italian varieties has been suggested by Merlo (1925: 238, 239), and supported by Bottiglioni (1926) who considered phonetical features common to Corsican, Sardinian, Sicilian and the central Italian dialects, and in particular consonants: "Il consonantismo còrso è insomma strettamente connesso con quello sardo–siculo e italiano centromeridionale" [The phonetics of Corsican consonants is related to the phonetics of Sardinian–Sicilian and central Italian; translation KAN] (Bottiglioni 1926: 64). This approach is also supported by Chiorboli (1988, 1991b: 19, 1992), who names the area the 'Romania intertyrrenique'. He lists features from different domains. For morphosyntax this concerns: the masculine plural ending in -a; the partitive construction based on the numeral *du, dui* 'two'; and polymorphic personal pronouns of the first person singular, all of which are common to Sicilian, Corsican and southern Italian varieties (Chiorboli 1992: 149, 151, 154). In a similar vein, see also Dalbera-Stefanaggi (1991b: 164, 165) and Fabellini (2010: 91, 92).

For the classification of Corsican among the Tuscan varieties, the focus has been on northern Corsican. By contrast, southern Corsican varieties are more closely related to the northern Sardinian variety Gallurese. In this tradition, Meyer-Lübke (1890) considers Gallurese as being part of the Corsican varieties and Bottiglioni (1957a [1936]), referring to lexical and phonetic features (identified by data collected for the *Atlante Linguistico Etnografico Italiano della Corsica*, ALEIC), postulates a historical "unita sardo-corsà" (Bottiglioni 1957a [1936]: 131) whereby the common linguistic variety has developed under the influence of Tuscan. Petkanov (1941) also follows this approach, considering phonetic features of Gallurese, and ascribes these features to a "phase ancient du corse meridional […] un corse pre-toscan" [ancient phase of medieval Corsican […], a pre-Tuscan Corsican; KAN] (Dalbera-Stefanaggi 1999: 167). While Petkanov classifies Gallurese as a Corsican(-based) variety, Lausberg (1963: 56 ff.) assigns Corsican, with Gallurese and Sassarese, to the northern Sardinian varieties – i.e. considers southern Corsican as a northern Sardinian variety.

Within the framework of the NALC-project, Dalbera-Stefanaggi also dedicates herself to the Corsican-Gallurese variety ("l'ensemble corse-gallurien" [The entirety of Corsican and Gallurese; KAN] Dalbera-Stefanaggi 1999: 161), and gives a detailed overview of the history of Corsican and Sardinian linguistics (Dalbera-Stefanaggi 1999: 161–171).

3.1.3 *'Original' Corsican*

In the context of diatopic differentiation and classification, the issue of 'original' Corsican must also

be mentioned, as being almost as old as the discipline of Corsican linguistics itself.[24] As long ago as 1916, Bertoni (1916) noted "il vero dialetto corso, parlato da circa due terzi della popolazione, è propriamente il cismontano, entro il quale si può distinguere il capo-corsino parlato nell'estremità settentrionale" [The true Corsican dialect, spoken by about two thirds of the population, is (the) *cismontano* (variety) within which we can discriminate *capo-corsino* at the extrem north; translation KAN] (Bertoni 1916: 147). Even if the concept of *lingua polinomica* (see below) is accepted, applied and integrated in everyday life, an idealistic idea of a *corsu sputicu, veru* 'pure, true Corsican' is still vivid, ignoring the influence French has on Corsican. Purists still believe that this "langue mythique" (Thiers 1986: 70; Marcellesi & Thiers 1988: 818) can be revived and actualised for modern reality. The discourse about 'original' Corsican reveals also the ideological value Corsican has, for example in the inclusion of certain authors and their work among 'original' Corsican writers and books, the ongoing discussions about 'pure, original Corsican' (see e.g. Bertoni 1916; Ettori 1975; Blackwood 2004b and many others), the significance Corsican has for the political autonomy endeavour (in particular in the 1970s and 1980s, cf. Melillo 1977: 1512), or the questions of which Corsican should be taught in school and the motivations for pupils to learn the language.[25]

Under the heading of language elaboration and language planning, Blackwood (2004b) also deals with this issue. He presents 'two Corsicans' as the result of the ongoing language planning policies. The first, labeled 'Gallicised Corsican', is characterised by the permanent language contact with French; this(these) variety(ies) fall within the scope of the *polynomia* concept (see below). The second, 'distanciated Corsican', is the result of distancing strategies from French within the language elaboration process; this variety is based on ideas of a *corsu sputicu*, as mentioned above (cf. Blackwood 2004b: 233). His classification dates partially from the "competing attitudes" described by Thiers (1993: 265–267), labeled therein 'Corsican by distanciation' and '*Corsité polynomique*'.

Thiers (1993), based on his work of the mid-1980s, presents two lines of language elaboration.

---

[24] The issue of what is (the) 'original' Corsican not only influenced language documentation during the past decades but is still important today, i.e. when conducting empirical fieldwork (for more details, section 5.4.3).
A first public discussion about the standardisation of Corsican can be found as soon as 1863/64 in the journal *Avenir de la Corse,* lead by the contributions of the Corsican Antoine Mattei. Although the selection process for a Corsican standard started with this discussion, it was not completed sixty years later when, in 1925, *A prima grammatichella corsa* of Antone Bonifacio was published. This grammar did not contain any codification of a Corscian standard, not did it even present a consistent spelling. The spelling varied until the 1960s. Even today, more than 150 years later, the selection process has not concluded but has been superseded by the application of the concept *lingua polinomica* (see below).

[25] The young(est) generation of Corsican speakers, being asked for their motivation to study Corsican (on a non-compulsory level), mention identity-based reasons: "the language chosen was that of the region, the family, the community culture. Those reasons outstripped other reasons suggested such as its usefulness in understanding other languages, or the value of being bilingual" (Fusina 2000: 23, based on a survey conducted at the secondary education level by the National Office of Information on Teaching and the Teaching Progressions [*Office National d'Information sur les Enseignements et les Professions,* ONISEP] in 1996).

The distantiation strategy, which "claims to be based resolutely on traditional oral use" and "seeks maximal differentiation from French" (Thiers 1993: 265), rejects even French-based vocabulary long-since integrated into Corsican, and substitutes it with 'purer' forms; e.g instead of *differenza* 'difference' or *aviò* 'airplane', we find *sfarenza* and *aeriu*. This approach "precludes contact phenomena, tolerates only minimal variation and is proclaimed as the rebirth of an older oral tradition, ignoring the historical reality of contact and conflict" (Thiers 1993: 265). Along similar lines, Blackwood (2004b) also argues for distantiated Corsican. By contrast, Gallicised or polynomic Corsican is instead a result of the language contact situation; existing ties between Corsican and French and Italian are actively strengthened, e.g. by Corsican neologisms based on French or Italian roots.

Hence, the concept of Corsican as a *lingua polinomica* matches the current linguistic reality better than any unidimensional account. Jean-Baptiste Marcellesi, a Corsican sociolinguist, argues as early as the 1980s for a *polinomica* concept, and explains it in the following way:

> Pour les langues romanes en train de se constituer ou de se faire reconnaître (en France: Corse, Occitan) cela conduit a proposer le concept de 'languages polynomiques'; c'est-á-dire de langues don't l'unité est abstraite et résulte d'un movement dialectique et non de la simple ossification d'une norme unique, et don't l'existence est fondée sur la décision massive de ceux qui la parlent de lui donner un nom particulier et de la déclarer autonome des autres langues reconnues. (Marcellesi 1984: 314)

> [For the Romance languages in the process of evolving or becoming recognised (in France: Corsican, Occitan) the concept of 'languages polynomiques' was proposed; the unit of these languages is abstract and results in a dialectic movement and not in a simple ossification of a unique norm. And, therefore, the existence is founded in the strong decision of the speakers to give them a specific name and to declare them independent from other recognised languages; KAN]

Originally (at the time of its conception), the *polinomica* concept applied primarily to the collectivity of Corsican diatopical varieties and referred *grosso modo* to the equal appreciation and recognition of all theses diatopic varieties. In practice, it results, for example, in the acceptance of all varieties spoken by children at school, despite the textbooks and the teaching being aligned with the two varieties *corsu supranu* and *corsu suttanu*. Following this strategy, grading in language classes according to traditional standards becomes difficult. But it assures that children acquiring Corsican

in their private surroundings keep their particular variety; and, in general, the transmission of the declining local (micro-)varieties, spoken by the children's families or members of the same village, is ensured. Also, the publishing house Albiana in Aiacciu recognises the individual author's variety and its characteristics as long as it is consistent within her/his work.

The purview of the *polinomia* concept widened gradually: not only is diatopical variation included, but so now are, for example, the language-contact induced varieties (at least theoretically). These varieties, which one could call 'intervening', 'intermediate', 'hybrid' or '*interlecte*', are the result of the ongoing exposure of speakers of Corsican to other varieties and languages. Ottavi (2011: 117) mentions the Corsican–French hybrids "*francitan, francorse, corsanceses…*".[26]

Even if the polynomic approach allows a fairly high degree of freedom in oral language use, in the written domain we find a phonographic orthography: certain phonetic units correspond to a certain spelling, i.e. grapheme(s);[27] no standardisation has been undertaken in other domains such as morphology, syntax or lexis (e.g. northern Corsican (*corsu supranu*) *cane* vs. southern Corsican (*corsu suttanu*) *ghjacaru* 'dog').

3.1.4 *Sources for studying Corsican and its history*

In spite of the implementation of Corsican as an independent field of study (e.g. the course *Corsican language, literature and culture* at the University of Corti) linguistic diachronic research on Corsican is still (and will continue to be) difficult to carry out. As Nesi (1992) states, the scarcity of data and the absence of a inventory of sources make diachronic linguistic analysis, which would need a quantitatively significant number of sources spread over a specific time-period and belonging to different types of text, extremely difficult. Even more so if the truthworthiness of a text has to be questioned because of possible changes during its transmission (cf. Nesi 1992: 922, Graziani 1929). Indeed, obtaining written or registered language data for linguistic research in places geographically distant from Corsica is still rather difficult. Muljačić (1989) mentions specific problems when it comes to original Corsican documents: "L'informazione bibliografica (specie quando si tratta di lavori ciclostilati) è incompleta, tarda parecchio e anche quando esiste non è di grande aiuto a studiosi

---

[26] Referring to sources for the terms of *francitan, francorse, corsancese*: Colombani (1968) introduced the term *francorsu* for the varieties based on language contact of Corsican with French. Pascal Marchetti, at the *Congrès international sur le bilinguisme* at Bastia in 1984, differentiates between *u francorsu* and *u corsancese* and defines the latter as "corse parlé à la française" (Thiers 1997: 1208). The attempt to classify the contact-based varieties (Corsican–French, Corsican–Italian) has been the object of various studies since the 1980s, e.g. Comiti (1992), Filippi (1992). A detailed overview of these varieties can be found with Thiers (1997) for Corsican–French, and (Chiorboli 1997) for Corsican–Italian.

[27] After decades of discussions, Marchetti & Geronimi (1971) with their book *Intricciate è cambiarine. Manuel pratique d'orthographe corse*, provided guidelines for the Corsican orthography widely accepted nowadays.

che vivono lontano dai centri in cui si insegna la linguistica, la filologia e la civiltà corsa" [The bibliographic information (in particular if it concerns hectographic work), is incomplete, takes a long time and even if there is any [information], it is not of great use to scientists who live distant from spaces where they teach Corsican linguistics, philology and cultural science; KAN] (Muljačić 1989: 287). Efforts made by the University of Corsica in Corti and the archives and libraries in Corsica, to review the existing inventory, do not go beyond the island's border, with the exception of some universities in France where Corsican is an integrated part of the listed degrees (For example, *Certificat de Langue, Littérature et Civilisation Corse* at the universities of Aix en Provence (founded in 1968; since 1975 with the *Centre d'Études Corse*), Nice (in 1970) and Paris III (in 1972)) (cf. Ettori 1975: 105; Janich & Greule 2002).

The situation described by Muljačić (1989) and later by Fabellini (2010) has not changed over the past twenty years. For example, even where Italian libraries house Corsican written documents, these cannot be found by means of the library catalogues; further, Corsican is not offered as a language-criterion for queries; nor, especially for older documents, may Corsican (book) titles be inserted in the search field, as Corsican titles have been Italianised either by the editor or the library itself.[28] For the availability of Corsican language documents of various kinds, the only geographically and institutionally independent source is the internet. But obtaining the earliest, and even some of the more recent original documents (as scanned or print sources or audio-registrations) is very difficult outside Corsica (a good overview about older documents written in Corsican and in Corsica, and literal reprints thereof, can be found in Nesi 1994).

These issues of language documentation and language sources are related to the issue of 'original' Corsican. The question of whether to include or exclude texts in linguistic analyses or even in a compilation of 'original' Corsican documents is often decided more on the basis of ideological ideas than analytic ones (for example, see Dalbera-Stefanaggi 1997: 303; Arrighi 2002: 7). Nesi (1992) demonstrates this with reference to the anthology compiled by Yvia-Croce (1987 [1929]):

Yvia-Croce nella sua raccolta pone al centro quell'essere corso – talora solo di natali – quale matrice culturale di scrittori che hanno trovato in lingue diverse il loro modulo espressivo [...] In generale è apparsa finora preminente l'attenzione per la testimonianza letteraria, per lo scrittore qualche rappresentante della forza spirituale di un popolo, mentre la veste linguistica dei testi nel suo valore di fonte documentaria desta minor

---

28 Beyond the basic issue of access to the documents, the unified Italianised graphics result in even more substantial problems: as Fabellini (2010) states for her diachronic studies, Italianised graphics in Corsican texts obscure phonetic-phonological but also morphosyntactic differences.

attenzione. (Nesi 1992: 923)

[In his anthology, Yvia-Croce gives priority to 'being Corsican' – at that time designated by birth – the cultural array of writers who have found in different languages their literary device […] So far, the attention for literature as a cultural medium hinges on the writer as a representative of a spiritual power of a people, while the linguistic aspect of their texts as a valuable documentary source gains less attention; KAN]

Even if Yvia-Croce's practice might be arguable from a linguistic point of view, it initiates the rethinking of the former approach, the classification of 'original, pure Corsican'. Considering the relatively small quantity of documents in general, a different point of view is needed, on the definition of the so-called 'original text', to open new possibilities for diachronic research on Corsican: that is, the recognition of the linguistic reality prevailing in Corsica for centuries; from the diglossic situation of written Italian and spoken Corsican before the arrival of the French; to the Corsican–Italian– French triglossia; until today, when French dominates every domain. These changes brought and bring about linguistic interferences which are part of the current linguistic reality. Therefore, we need a non-restrictive (or at least less restrictive) practice with Corsican documents (see already Chiorboli 1978: 156; Nesi 1992). This may also apply to dealing with linguistic informants and their language competence (see Chapter 5).

With a less restrictive approach, we may analyse, for example, written Italian in Corsica by Corsicans, and look for Corsican traces:

[…] una posizione non restrittiva nei confronti delle testimonianze disponibili offre la possibilità di un assaggio di già che è stato l'italiano scritto di Corsica nelle sue diverse realizzazioni, indubbiamente legate ai diversi utenti ed anche al di là delle intenzioni letterarie degli autori (Nesi 1992: 923).

[…] a non-restrictive position regarding the available sources offers the possibility of a sample of Italian written in Corsica and its different realisations, related to its different users and thus to the different intentions of the authors; KAN]

Against the backdrop of a less restrictive practice in dealing with Corsican written sources, more possibilities for diachronic studies are offered by documents, albeit of limited length, found in the

Gallura region from the 15th century onwards.[29] That implies that Corsican was not only an oral but also a written variety at that time. Beside the important observation of Corsican being used in written contexts, Maxia (1999: 36) claims that what he calls the "còrso illustre" is very similar to written Italian.[30]

Maxia's (1999) statement is interesting with respect to Nesi's (1992) idea of a progressive assimilation of Corsican to Tuscan Italian from the 18th century onwards. Any local documentation in Corsican, up until the 17th century, consisted of chronicles, letters and statutes kept in the ecclesiastical archives and recorded in copies from the 18th century (for detailed information, see Nesi 1992: 923 ff.; 1994: 895–911). Only starting from the 18th century, we find more written documents. An along with these ducuments, we can also see an increasing assimilation to Tuscan Italian for literary texts (cf. Nesi 1992: 924); fewer local or regional interferences can be detected

---

[29] For the *periodo giudicale* (between 900 and 1420), toponyms based on Corsican names, ethnicities, or other vocabulary, are attested in the Sassari region. (cf. Maxia 2006: 83)

Three letters written in Corsican: *Deposizione del rettore della chiesa di S. Niccolò di Spano* (20.09.1400); *Lettera di prete Polino da Mela ai protettori del Banco di San Giorgio* (11.06.1498); *Lettera dall'esilio di Giovanpaolo Leca, conte di Cirnarca, ai figli* (written in Sassari, 02.06.1506) (cf. Maxia 1999: 35).

The oldest testimony of Gallurese is a collection of poetries, dating back to the beginning of the 18th century. The poetry is written by Gavino Pes, also known as *Don Baígnu*, a religious man who lived in the 18th century in Tempio (cf. Corda 1983: 5). Maxia (1999: 105, 106) claims that Pes' Gallurese, as far as its phonetics, morphology, syntax, and lexis are concerned, does not diverge from present-day Gallurese of Tempio except for some Spanish lexemes not in use anymore. Hence, present-day Tempio-Gallurese is very similar to how it was in the 18th century.

Nevertheless, as for the centuries before, all demographic documentation (i.e. certificates of baptism, marriage, and death) archived in the parishes of Aggius, Rampio and Calangianus was still written in Logudorese (cf. Mameli 1998: 139). However, despite the well-preserved documentation of the Corsican immigration, these documents do not reveal any detailed information about linguistic developments in their respective regions.

[30] Citation: "Ciò significa, dunque, che il còrso non soltanto era parlato anche dagli strati sociali più elevati ma conosceva un uso scritto sia pure limitato, a causa dei pochi dati di cui siamo in possesso, ai rapporti fra Sassari e la Corsica. Abbiamo visto, inoltre, dalle lettere del Cano e del conte di Cinarca, che il còrso 'illustre' non differiva in modo particolarmente significativo dall'italiano attestato sulla penisola durante il medesimo periodo. Le differenze, in effetti sembrano circoscritte a pochi fatti lessicali, all'accentazione, al particolare vocalismo, a qualche aspetto del consonantismo e, soprattutto, alle forme verbali" [This means, that not only was Corsican spoken by the higher social strata but that it was also used in writing, even if this was limited to relations between Sassari and Corsican, according to the limited data we have. We have seen, moreover, from the letters of Cano and the Count of Cinarca, that 'illustrious' Corsican did not differ in any significant way from the Italian spoken on the the peninsula during the same period. The differences, in fact, seem to be limited to a few lexical facts, accentuation, particular vocalism, come aspects of consonantism, and, above all, verb forms; KAN] (Maxia 1999: 36).

from then on. In the few preserved non-literary texts we find only traces of spoken Corsican.[31]

In the 19th century, the elaboration (*Ausbau* in the sense of Kloss 1978; Bossong 2008: 25 ff. for the Romance languages) of written Corsican starts in the domain of elaborated orality that is songs, sayings and tales. The *Ausbau* process in Corsican, its "Verschriftlichung" [textualisation; KAN] (according to Oesterreicher 1993) does not occur in an "Umgebung der primären Oralität" [atmosphere of primary orality; KAN] (Fabellini 2010: 99). Corsicans, in particular those devoted to literary work, are already competent writers of Italian and/or French. For more detail on the Corsican *Ausbau*, see section 3.2.3 and Table 10).

According to Bossong (2008), the elaboration of a language variety is a process, and therefore a continuum. Reference points within this continuum can be literary genres and topics of expository prose. Usually, "Poesie ist schon bei sehr geringem Ausbau präsent, es folgt Theater und erst ganz am Ende narrative Prosa" [Poetry is present at an early stage of *Ausbau*, followed by theatre and at the very end, narrative prose; KAN] (Bossong 2008: 26). Beyond objective criteria, the speakers' attitudes towards their language variety are also important, in combination with the political situation; for more detail see Bossong (2008: 25–28).

The *Sirinata di Scappinu*, written in 1817 and published in 1823 is commonly considered as the first literary text written in Corsican. It constitutes one part in the otherwise Italian epic *Dionomachia* (a "poemetto eroi-comico", Viale 1817: 3) by the Corsican author Salvator Viale. Most Corsican texts of the 19th century are embedded in an Italian or French main text. For Fabellini (2010) this fact serves as proof that "[d]as Transponieren des Korsischen in die Schriftlichkeit […] zu diesem Zeitpunkt noch kein Zeichen dafür [ist], dass dem korsischen Idiom kommunikative Eigenständigkeit zugestanden wird" [The transfer of Corsican into the written domain at the time is not yet a sign of Corsican gaining communicative independency; KAN] (Fabellini 2010: 106). But already by the second half of the 19th century, Santu Casanova (1850–1936), also called *Ziu Santu* – an Italian-writing poet, one of the most important writers of Corsican popular literature, and the later founder

---

[31] Citation: "Più scarse […] le scritture non letterarie che abbiamo raggiunto pre questi secoli. Scarsi dialettismi o localismi linguistici emergono dalla corrispondenza esaminata, soprattutto, come è ovvio, se è relativa a rapporti istituzionali o se si tratta di scambi fra letterari […] per i primi anni del XIX secolo, l'opera manoscritta di Giuseppe Maria Saliceti, medico a Loreto di Casinca, laureato a Roma, nella quale espone e discute i casi clinici in una prosa snella, non sempre sintatticamente chiara nella quale si affianca ai termini e alle locuzioni tipiche della scienza medica di quel tempo un elevato numero di elementi del parlato […] la frequenza di consonanti sonore […] o i casi di ipercorrettismo" [More scarce […] are the non-literary writingd recorded for these centuries. Few dialectisms or linguistic localisms emerge from the correspondence examined, above all, as is obvious, if it relates to institutional relations or if it is an exchange between literaries […] for the early 19th century, the manuscript work of Giuseppe Maria Saliceti, a doctor in Loreto di Casinca, graduated in Rome, in which he expounds and discusses clinical cases in a stramlines prose, not always syntactically clear, in which the terms and locution typical of the medical science of the time are flanked by a large number of spoken elements […] the frequency of sound consonants […] or cases of hypercorrection; KAN] (Loi Corvetto & Nesi 1993: 233, 234).

of the magazine *A Tramuntana* – was convinced that Corsican can be used for all kinds of topics. Besides translating Italian texts into Corsican, he wrote independent works of different genres in Corsican, such as the burlesque *Testamentu di Francescu* (1875) and the collection of essays *Primavera Corsa* (1927). Thus, at the end of the 19th century, the first independent literary Corsican works arose, including, among others, the poems and comedies written between 1888 and 1903 by Petru Lucciana, followed by the first tragedies and dramas around the turn of the century. The genre novel was discovered late; the first novel was published in 1930, *Pesciu Anguilla,* written by Sebastiano Dalzeto.

Already, from 1863 onwards, the articles of Antoine Mattei, published in the newspaper *Avenir de la Corse*, opened the metalinguistic discourse on written use of Corsican to a broader public. Mattei was in favour of a structured elaboration. He undertook preparatory work with the collection of documents, drafts for dictionaries and works on onomastics which he completed by means of comparative language studies. The metalinguistic discussion was carried on in (literary) journals founded at the same time, such as the aforementioned *A Tramuntana,* of which the first volume was published in 1896. Also for the first time, social and political topics were discussed in written Corsican.[32] The *Vocabolario dei Dialetti della Corsica* by Francesco Domenico Falcucci (1835–1902), published posthumously in 1915, presented the first lexicographic work on Corsican and animated the discussion about the Corsican orthography during the first half of the 20th century.

The literary use of Corsican progressed quickly. The most important decades are those between the 1870s and the 1920s. According to Piazza (2009), the impetus for the *Ausbau* from oral to written Corsican came from the linguistic rearrangement, the supersession of Italian by French as an instrument of literary production. Corsican writers were in crisis: "C'est la crise d'identité provoque par la substitution du français à l'italien que va provoquer la transformation de la littérature corse orale en littérature écrite d'expression corse […]" [This is an identity crisis caused by the replacement of Italian by French and which will provoke the transformation of the Corsican oral literature to written literature; KAN] (Piazza 2009: 196). After a slow and gradual transition, Italian was officially substituted by French within the 19th century (cf. Piazza 2009: 191).

Beginning with the reconstitution of the *Société des Sciences* in 1948, which had ceased its activities from 1937 and throughout WWII, the 1950s were characterised by a spirit of optimism regarding promotion of Corsican on different levels of cultural life: some authors of Corsican origin produced parts of their literary work in Corsican, such as Marie Susini with *La Fiera* (in 1954); the

---

[32] For more about the elaboration process of Corsican in the literary context and its theoretical embedding into the Kloss's *Ausbau* model see Fabellini (2010: 102–108).

tradition of the (monthly/bimestrial/etc.) journal was continued, e.g. by Pierre Ciavatti who founded *Le Muntese* in 1955; and a first version of the *Études Corses* was published together with other journals by the *Société des Siences*. A first attempt to bring Corsican into the world of university studies was conducted by Paul Arrighi, founding the *Centre d'Études Corses* at the university of Aix-en-Provence (for more detail, see Piazza 2009: 206, 207).

All these activities cannot hide the fact that it was mainly the older generation of authors and activists, born before WWI, who were involved in the process; the younger ones, born between the two wars, were not part of it, with few exceptions. So the thriving movement of written Corsican, reaching a first peak in the first decade of the 20th century, was not enhanced by younger followers. Already in the 1950s the intergenerational gap was noticeable, even if somewhat compensated by the activities of the elders:

> Le bilan que l'on peut tirer de la période que court de 1945 à 1975 est mitigé. On notera d'abord la prédominance de la production poétique, mais l'impression de quantité est trompeuse. La plupart de ces poètes sont des survivants de la génération d'avant 1914 […] La génération de l'entre-deux-guerres ne compte pratiquement plus de Corses parlant le corse comme langue maternelle parmi les écrivains […] Pour la littérature en prose, la situation est la même et les anciens masquent le vide de la générations […] (Piazza 2009: 207).

> [The result of the period between 1945 and 1975 is mixed. Above all, one notices the dominance of poetry, but the quantitative impression is misleading. The majority of these poets are survivors of the generation before 1914 […] The generation of the interwar period does not count as the writers do not speak Corsican as a mother tongue among themselves […] For prose, the situation is the the same and the elders hide the gap between the generation; KAN]

From 1970 onwards, Corsican literature gained new vitality from authors born since 1930, for example, Dumenicantone Geronimi, Rinatu Coti, Ghjuvan Ghjaseppiu Franchi, and Santu Casta – authors of all literary genres. New journals were published, e.g. *Rigiru*, founded in 1974 by Dumenicantone Geronimi. Also, theatre groups and music groups were established and a broader public, even beyond the Corsican borders, rediscovered the traditional polyphonic music.

During the past fifty years, the *Ausbau* process has expanded into more and different domains of public life – besides the literary genres, functional texts such as conference publications,

documents and forms, public advice. From that point, the active *Ausbau* of Corsican vocabulary has gone in different directions: some favour an Italo-Romance basis in combination with Corsican morphology; others chose words that display a certain proximity to French, to underline Corsican's independence from Italian (see the aforementioned comments on polynomic and distantiated Corsican in section 3.1.3).

3.1.5 *State of the art: linguistic studies based on empirical data*

The lack of reliable sources of Corsican written texts also affects empirical linguistic data; even if most authors of studies on Corsican do elicit Corsican language data, the sources of the data are often not (explicitly) revealed (e.g. Casta 2003), or their deficient citation is nowadays attested (e.g. Rohlfs 1941). Therefore, the compilation of empirical studies presented below only includes studies where the origin and compilation of the data have been documented. Due to the size constraints for this thesis, I present only a selection of studies (from different linguistic domains).

Documentation of Corsican based on oral data has a vivid tradition in the form of linguistic atlases. Three linguistic atlas projects have been launched since the beginning of the 20th century: in the early 20th century, the *Atlas Linguistique de la France: Corse* (ALF C) by Gilliéron & Edmont (1914) and the ALEIC by Bottiglioni (1933–1942); and, since the mid 1970s, the NALC-project.

In 1914, Jules Guilliéron and Edmond Edmont published the ALF C. For their work, Edmont, not being Corsophone, studied Italian before he conducted the survey, posing about 3,000 questions in 44 localities in 1911 and 1912 (cf. Dalbera-Stefanaggi 2002: 64). The project was stopped before completion by the outbreak of WWI. As the ALF C consists of single lexical items, it does not permit syntactic analyses.

A broad collection of language data is offered by the linguistic atlas ALEIC, written by Gino Bottiglioni. He collected data in 49 locations on the island, completed by 2 locations in northern Sardinia, 1 on Elba and 3 in the western Tuscan area. The survey was based on 1,950 sentences and resulted in 2,000 maps, published in 10 volumes between 1933 and 1942. The maps are completed by a dictionary, listing all vocabulary items collected during the survey, plus an introductory work explaining the methodology and the theoretical background. Much more extensive than the ALF C, Bottiglioni's atlas is based not on isolated lexical items, but on idioms (i.e. *che puzzo!* 'what a stink!', Map 116) and sentences (i.e. *preghiamo Gesú Cristo e la Madonna!* Map 1904; *volete che ci vada, o che ci mandi qualcuno?* Map 361). Bottiglioni was aware that a survey based on sentences would also be beneficial for further analyses. His primary interest was the diatopical variation and the classification of varieties on the basis of phonetic and lexical characteristics. Although the ALEIC

permits syntactic analyses (as also conducted by its author, cf. e.g. Bottiglioni 1957c [1932]), it only contains simple matrix clauses, simple interrogative clauses and idiomatic expressions, but almost no cases of complement clauses or complex subordination. Additionally, the context of the respective sentences is missing, which reduces the possibility of research on context-sensitive phenomena such as discourse prominency – a crucial trigger for DOM in Romanian (cf. von Heusinger & Chiriacescu 2011).

For any synchronic language study, both atlases should be used with caution, as the Corsican language varieties and their use have undergone major changes, especially within the 20th century. Speakers of Corsican, especially the generation born after the two world wars, have acquired their language skills under different conditions to the generations growing up before the wars; among other reasons, language transmission was reduced by the loss of male native speakers due to the wars, the working environment changed (e.g. the decline of traditional fields of works and emigration), and higher education, which became increasingly important, was provided only in French and on the (French) mainland. These circumstances also changed private language use.

At the moment, the NALC is the most current, large-scale approach to language documentation which allows for synchronic language studies. The NALC started as part of the larger project *Atlas Linguistiques de la France par Régions* in 1974, under the direction of Mathée Giacomo-Marcellesi.[33] Since 1981 Marie José Dalbera-Stefanaggi has directed the Corsican atlas at the university of Corti. In 1986 the NALC-project was extended to include a database, the *Banque de Données Langue Corse* (BDLC). Since then the NALC–BDLC project has been launched under the triple auspices of the *Centre National de la Recherche Scientifique* (CNRS) (for NALC), the *Collectivité Territoriale de Corse* (CTC), the university of Corti (for BDLC).

When taking over the project in 1981, Dalbera-Stefanaggi and team were conscious of the need to adjust the project, including media development and methodology, according to advances in technology and developments in linguistic science, while preserving material already gathered (cf. Dalbera-Stefanaggi 1989: 137). Being aware that a traditional linguistic atlas – even if developed according to the guidelines of dialectology, i.e. "dimension spatial, dimension temporelle, dimension sociale et dimension culturelle" [the spatial dimension, the temporal dimension, the social dimension and the cultural dimension; KAN] (Dalbera-Stefanaggi 1989: 138) – would not be sufficient for future research in different linguistic domains, they began the development of a computerised

---

[33] Prior to the surveys for the NALC, Giacomo-Marcellesi and colleagues had already conducted surveys on single varieties of Corsican between 1969 and 1975 (cf. Giacomo-Marcellesi 1978b: 140). As a continuation of these research activities, a preliminary questionnaire for the NALC was also developed. This questionnaire was based on the one used for the *Atlas linguistique du Lyonnais*. It contained 1,219 questions, including descriptions of the collected lexemes.

database, the BDLC, in 1986. The database constitutes not only raw language data, i.e. words and texts including also recorded audio-material, but also analytical information ("données brutes et analyses", Dalbera-Stefanaggi 1995: 161). In the three printed NALC-volumes the data is presented under its lemma, in phonetic transcription using the International Phonetic Alphabet (IPA), and displayed in a cartographic illustration.

BDLC contains different kinds of information – 'raw' data (e.g. audio-files) and processed (e.g. annotated) data – to make different levels of information available to users of the database. The processed data include, for each entry, the phonetic transcription, a lemma, an etymological form, morphological structuring, and additional ethnolinguistic information, i.e. about iconography and bibliographical references.[34]

Studies published by Dalbera-Stefanaggi (1989; 1995; 1997; 2001; 2002; 2004; 2007a)[35] are all related to the NALC-project and based on data gathered for the same.

From the late 1970s and during the 1980s, the official status of Corsican became increasingly stable due to, for example, the reopening of the university in Corti, the academic course on Corsican language and culture, and the implementation of structured language teaching in public education. At the same time a seemingly unstoppable decline of language practice in everyday life was observed (cf. Thiers 1986: 65). This was a time when linguistic empirical research on Corsican increased; in particular, in the domain of sociolinguistics, the elicitation and description of synchronic empirical data gained centre stage. The phonetic and lexical domains retain constant attention. The studies that are presented below predominantly have a dual alignment: they include linguistic and metalinguistic data, analysed both quantitatively and qualitatively.

Jean-Baptiste Marcellesi, who is particularly associated with sociolinguistics and the implementation of Corsican in (public) education, has, with regard to language teaching, also been concerned with

---

[34] The technical structure of the BDLC is based on different source files: (1) The basic file (which links the item with the file of locations), the question-file (which included the respective 'question' of the questionnaire), keywords (for the thematic fields to which the entry belongs), and bibliographical references; (2) files with etymological, morphological and lemmatic information, are linked to the basic file (cf. Dalbera-Stefanaggi 1989: 144). This structure allows different levels of usage: a kind of French–Corsican and Corsican–French dictionary based on the lemmata, while also making available the etymological entry; (3) another level, also based on the lemmata, displays diatopical variation, and makes available a cartographic representation of the entries; (4) finally, the existens of 'tagged' information allows for linguistic (and other) analyses to be performed (cf. Dalbera-Stefanaggi 1989: 145, 146)

[35] The work on Corsican of Dalbera-Stefanaggi can be considered as the most extensive and – even if concentrated on phonetics/phonology and lexis – the most multifaceted on the topic. Thiers (1986) states: "À ce jour la description la plus complète est l'ouvrage de Marie-José Dalbera-Stefanaggi" [The most complete description to date is the work of Marie-José Dalbera-Stefanaggi; KAN].

the '*indicateurs d'identité*' (Marcellesi 1983: 399).

In the mid-1980s, Ghjacumu Thiers and colleagues[36] dedicated themselves to the issue of '*Implication théoriques et modalités du processus d'individuation sociolinguistique corse*' in terms of a project financed by the CNRS and located at the university of Rouen. In the context of this project an epilinguistic study was also conducted in collaboration with the students of the study course *Études corses (option sociolinguistique)* at the university at Corti (cf. Thiers 1987). Thiers describes the goal of the project in the following terms: "Le but que se propose le chercheur n'est pas de décrire le ou les dialectes d'un espace géographique donné, mais d'étudier comment des locuteurs appartenant à un espace géographique et social donné prennent conscience des traits individuateurs de leur(s) dialecte(s) […]" [The goal is not to describe the dialect or the dialects of a given geographical area, but to study how the speakers, who belong to a certain geographic and social space, become aware of the individualising features of their dialect(s); KAN] (Thiers 1987: 44). The results of the survey should also be considered in didactic terms: "Les recherches de l'équipe trouvent aussi leur application dans le domaine de la didactique du corse. Sur ce point est venue confirmer et amplifier un mouvement que n'est pas nouveau: le souci d'établir l'enseignement du corse sur des principes variationnistes […]" [The investigations will also be applied in the didactics of Corsican. At this point (the project) confirms and reinforces known circumstances: the difficulties of establishing a method of teaching Corsican based on principles of variation; KAN] (Thiers 1987: 47).

For the survey they developed a collection of audio-texts in different Corsican and non-Corsican language varieties.[37] These audio-texts are based on a common French source text which was designed to include in its Corsican translation a high number of the so-called '*indicateurs de corsité*' (Chiorboli 1987: 77), features that are considered characteristic for Corsican. These characteristics refer to different linguistic domains: lexis, prosody, phonetics and phonology, syntax and morphosyntax, but are not further specified by Thiers (1987: 41, 42). Chiorboli, being a member of the research team, mentions adverbs, the vocative, prepositions, future tense, DOM, polysemous *chi*, subordination and the article system (cf. Chiorboli 1987: 77)

Regarding the *indicateurs*, informants mentioned lexical differences first, followed by prosodic differences (called 'musica' or 'accentu'). The informants made very few comments on intonation

---

[36] Among others, M. Giacomo-Marcellesi, J, Chiorboli and J.-B. Marcellesi.
[37] Thirteen varieties had been included in the first run in the following order: (0) French source text), (1) standard Italian, (2) variety of Bastia, (3) variety of Tavaru, (4) falsified variety with mixed features from northern and southern Corsican, (5) variety of Casinca, (6) variety of Aiacciu, (7) falsified variety without any regional features, (8) variety of the Balagna, (9) variety of Sartène, (10) variety of Sassari, Sardinia, (11) variety of the Gallura, (12) variety of Niolu, and (13) variety of Bunifaziu (cf. Thiers 1987: 42).

or the pitch level. The phonetic features were the most triggering: not only did informants identify the diatopical variety/varieties, but also the (metalinguistic) discourse about these features was most productive. Syntactic and morphosyntactic features were often recognised, but considered less important ("Les traits syntaxiques et morpho-syntaxiques […] apparaissent souvent […] mais tous ne sont pas également obligatoires et pertinents" [The traces of syntax and morphosyntax appear often, but not all of them are equally obligary and pertinent ; KAN] (Thiers 1987: 51).

In summary, it can be stated that speakers of Corsican, despite a general inter-comprehensibility among speakers of different varieties, are quite aware of diatopic variation:

> The islanders' perception in the centre of Corsican linguistic space, is of a basin where isoglosses merge into one another and micro-regional identity markers intertwine […] The distribution [of the features] has clear reference, in the linguistic consciousness of Corsican speakers, to determinate dialectal areas. (Thiers 1993: 259)

Besides the linguistic features under investigation, Thiers (1987) refers also to the behaviour of informants during the survey. This confirms a strong relationship between linguistic features and linguistic self-definition of the speakers:

> Les procédés de description et les arguments engagés par les P.I. [informants] éclairent en partie, dans ces opérations, l'influence que les indicateurs purement linguistiques exercent, par l'entremise d'une conscience linguistique active, sur la forme et les productions de l'imaginaire collectif interrogé sur la question de la corsité linguistique. C'est en somme la transformation du trait linguistique en fait sociolinguistique […] (Thiers 1987: 52).

> [The way to describe and and the arguments quoted by the informants explain the influence exerted by purely linguistic indicators in these processes, by an active linguistic awareness of the appearance and the production of a collective image of the linguistic *corsité*. This is the transformation of linguistic into sociolinguistic features…; KAN]

On the basis of the same survey, Chiorboli (1987) illustrates the diatopical variation of the main features in morphology and nominal syntax. He also considers the synchronic data against the backdrop of language change, and evaluates them against the process of language elaboration.

The British sociolinguist Robert Blackwood is interested in language policy and planning of regional

languages, language in the public domain and linguistic landscaping, in particular regarding Corsican.

Blackwood's (2004b) account on Corsican language planning strategies is based on two surveys conducted in 1999 and 2000. Setting aside the theoretical implication of the issue of language policy and its consequences for language elaboration (see section 3.1.3), the linguistic interest of these surveys concerns "the penetration of new distantiated vocabulary into the lexicon of the population [...]" (Blackwood 2004b: 237). The surveys were based on a French wordlist, for which Corsophones were asked to give the/a Corsican translation(s).

Only some of Blackwood's informants were aware of the effects of the ongoing language contact with French: they were able to classify their answers as French-based neologisms or lexical borrowings. But they could not explain their answers in terms of active language planning.[38]

On the basis of statistical analysis performed on the data gathered by the two surveys in 1999 and 2000, Blackwood concludes that the distantiation strategy is of secondary importance for the language revitalisation process, as the majority of the answers were Gallicised Corsican or Corsicanised French lexemes. Correlations between the linguistic data and informants' ages reveal that the average age of the Corsicanised French speaker was 61 (about 20% of all informants), while the average age of the distantiated Corsican was 39 (about 14% of all informants). The remaining 66% of the informants were not classified, as their answers did not display a consistent pattern.

In his book on language policy in Corsica, Blackwood (2008) compiled the results of five studies conducted between 1999 and 2005, including the two surveys already presented. The goal was to reveal the relationship between the state, language activists and the island's population with regard to language policy. The surveys were guided by the following definition of language policy: "language policy comprising of (*sic*) language management, language beliefs and language practices" (Blackwood 2008: 103, following Spolsky 2004)

The five surveys were similar in procedure and results. For the two questions "Do you understand Corsican?" and "Do you speak Corsican?" answers were self-reported and were tested. Based on these answers, the result for auding was almost the same for both genders, slightly higher for male informants than female. By contrast, speaking could be attributed unambiguously to male informants: 76% of the male informants, versus 24% of the female informants, spoke Corsican in the 2000 survey (cf. Blackwood 2008: 107). A total of 60% of the adult informants reported that they could speak Corsican.

---

[38] Only one of his informants (R12/99; teacher and language activist) was aware of the purpose of Blackwood's survey, and only after answering the last of the three questions.

On the other hand, when asked if they spoke Corsican with their family ("the L domain par excellence of the home" Blackwood 2008: 110) the total percentage of positive answers amounted to just 57% of informants. Blackwood explains this result in terms of the social behaviour of men: "men socialising together in the numerous bars and cafés of the island or playing boules together on the squares […] speak Corsican with friends and acquaintances rather than with blood relatives" Blackwood (2008: 110).

When specifying the family members to whom they speak Corsican, most informants reported that they spoke Corsican with their parents (32% of all informants) but only about 7% spoke Corsican with their children (cf. Blackwood 2008: 110, Table 7.4). Dividing the latter according to gender, we find 41% of men and 59% of women speaking Corsican with their children. Even though the total number is very small, it can be seen that women play an important role for intergenerational transmission.

Revently, other linguistic phenomena beyond diatopy and speaker's attitude come to the fore. A small study about Corsican prosody has been presented and published by Boula de Mareüil et al. (2012). This study is noteanle as being the first study on Corsican prosody. The goal of the study is to investigate "whether a prosodic transfer can be highlighted from Corsican […] to French spoken in Corsica […]" (Boula de Mareüil et al. 2012: 1). For the corpus, four bilingual Corsican speakers from the region around Corti were interviewed and audio-recorded. All the Corsican informants (two women, aged 50 and 72; two men, aged 35 and 57) were "very committed in the cultural and linguistic field" (Boula de Mareüil et al. 2012: 2). The data collected in Corsica were compared with the linguistic production of Parisian reference speakers, selected to match in age and gender. The corpus consists of data gathered via three different approaches: reading 60 sentences, declaratives and interrogatives; reading the fable "The North Wind and the Sun" and translating it into Corsican; and participating in a semi-directed interview in Corsican and French.

The results of the comparison show that, with regard to proparoxytone words (i.e. stressed on the antepenultimate syllable), no prosodic transfer to French occurs. But "a Corsican accent is especially perceived in [yes/no] questions" (Boula de Mareüil et al. 2012: 2).

This section has given an overview of studies based on empirical data. As mentioned before, dialectal studies as conducted by Dalbera-Stefanaggi and colleagues for the NALC dominate linguistic research on Corsican; phonetics, phonology and lexis have received particular attention. Phenomena relating to other fields of linguistics, such as morphology or syntax, receive less attention; if considered at all, many features are investigated together and presented at once (e.g. Chiorboli 1987),

with one or two phenomena sometimes being singled out (Marcellesi 1983). These morphological and syntactical features are predominantly listed among features of other linguistic branches (e.g. studies on the specific characteristics/linguistic indicators of Corsican, e.g. Thiers 1987; Nesi 1992). The metalinguistic discourse about Corsican in the context of language elaboration, language planning and language policy has led to the work of Blackwood (2004b; 2008), who recently widened his methodology by applying the linguistic landscaping approach (Blackwood 2014). Most recently, the field of prosody has also received attention (Boula de Mareüil et al. 2012).

3.2 **Historical background**

This section retraces the historical developments in Corsica which can be considered crucial to understanding the linguistic development of Corsican. Some linguistic traces of the pre-Tuscan time (until the 12th century) are still detectable. During the following Tuscan period, from the 12th century until the 18th century, Corsican was the spoken language, while Tuscan Italian served for written purposes. Hence, Corsica was *überdacht* ('roofed') by the homo-genetic variety of Italian. Corsica existed with this stable diglossia for more than five centuries. Since 1768, Corsica has been governed by France. As a consequence, French gradually occupied the previously Italian domains. This change implies a new *Überdachung* by the heterogenetic language, French. The stable diglossia has changed over time: moving from a diglossic situation, in which Corsican dominated the private spoken domains while Italian and French 'shared' the written domains; to the complete supersession of Italian by French.

3.2.1 *Pre-Tuscan Times: until 1120 AD*

Traces of human life can be found on Corsica dating from 7000 BC onwards, the population immigrating particularly from Liguria and Asia Minor. Traces of larger groups are found in the regions of Bunifaziu and Currachjaghju (to the north-east of Sartène). Only in the late Neolithic (6000–4000 BC, roughly) did the population start to spread over the island. Around 5000 BC, a steady influx of different groups of people introduced a multicultural population into Corsica. Traces of Tuscan-style pottery and tools made from Sardinian stone show a permanent interchange with the Italian mainland and Sardinia. Around 2000 BC a similar civilisation to that of Sardinia existed in Corsica, characterised by the Nuragic culture and a common Corsican–Sardinian pre-Latin substrate (cf. Arrighi 2002: 12). Then, from the 8th century BC onwards the Greeks occupied Corsica, founding Aleria (previously Alalia) in around 560 BC. Aleria became an important port in the Mediterranean Sea and was a vivid melting pot of cultures and nations. At the time of the Romans' conquest of the coastal regions (from the 3rd century BC onwards), there were Greek, Phoenician, Etruscan, Syracusan and Punic substrates. The inner parts of the island were mainly populated by the indigenous population (cf. Tagliavini 1972 [1949]). During the 1st century AD, Vandals invaded Corsica, as well as other parts of the Roman empire. The Vandals were later followed by the Lombards. Already during that time, the Moorish carried out offensives against Corsica. Around the year 1000 AD, the north was dominated by lords of Tuscan origin and the south by the *Cinarchesi*, descendents of *seigneurs* of the south (cf. Ottaviani 2010: 357) whose origins remain uncertain (cf. Arrighi 2002: 16). The geographic centre of the island remained unconquered for a long time.

Linguistic traces from the pre-Latin period can be found among Corsican toponyms (cf. Dalbera-Stefanaggi 2002: 7) and among isolated lexemes, which cannot be found in other parts of the Romance language family except for Sardinia. These lexemes relate the two islands to a common past (cf. Nesi 1992: 920).

The linguistic similarities on a lexical level can be seen in examples (6) and (7), the latter of which also reveals their common Latin origins:

(6)   Sard. *mènte,* South.Cors. *mènti*   'temple [body part]'
        (Rohlfs 1941: 14)

(7)   Lat. CIRCINARE > Cors. *circinà,* Sard. *chirchinnare*   'to cut in circular form'
        (Rohlfs 1941: 14)

From the Latin traces found in Corsican, Nesi (1992) derives a close relationship between Corsica and the Italian peninsula. Beyond purely linguistic traces, other historical facts strengthen this estimation. For example,Corsica played an important role in the missionary operation: "La cristianizzazione con i limiti degli apporti documentari si prospetta dunque come un periodo di forte e continuato legame con la penisola e soprattutto con l'area centrale, fino a determinare i rapporti con la Toscana che culminano nel dominio pisano" [The Christianisation despite its limited documentation appears as a period of a strong and persistent relationship with the Peninsula (i.e. Italy) and particularily with the central area, and determine the relationship with Tuscany, culminating in the Pisan sovereignty; KAN] (Nesi 1992: 920).

Cartain well-known linguistic phenomen, which are found in Tuscan varieties and attributed to a pre-Latin or early-Latin period, do not occur in Corsican. For example, Corsican keeps the final -*u,* as in Latin (see example 8) and does not show diphthongisation, as Italian does (example 9):

(8)   Lat. LUPUS, MANIBUS > Cors. *u lupu, a manu*   'the wolf', 'the hand'

(9)   Lat. PEDEM > Ital./Tusc. *piede* vs. Cors. *pede*   'foot'

Instead we find, as in some Southern Italian varieties and in Rome:
(i) the phenomenon of '*sandhi*' or consonant mutation (*scunsunatura*), the sonorisation of plosives. The pronunciation of consonants varies according to the preceding vowels: accented vowels cause a voiceless pronunciation of the word-initial plosive, whereas unaccented vowels cause lenition of the plosive, i.e. a voiced pronunciation (cf. Arrighi 2002: 34). This phono-syntactic

phenomenon is more marked in the north of Corsica than in the south.

(10)   *a porta*   'the door'   [aborta]   north

                                                                        [aporta]   south

                                                                        (Thiers 1993: 254; Arrighi 2002: 34)

(ii) the neutralisation of the vowel quantity, typical for Corsican:

(11)   Lat. PĬLU > Cors. [p'ilu]   'thread'

         Lat. FĪLU > Cors. [f'ilu]   'son'

Moreover, certain morphosyntactic phenomena in Corsican, such as the accusative–dative order of clitics (see examples 12 and 13) cannot be attributed to Tuscan. Only in the antique Fiorentino and the eastern Tuscan varieties can this be observed, while in Siena, Lucca and Pisa, only the reversed order (i.e. dative–accusative) can be found, since written documentation began. Loporcaro (2009) interprets it as a conservation strategy: "Se Pisa, fin dal Duecento, è passata al tipo innovativo, la presenza dell'ordine *lo+mi* nel còrso si può probabilmente attribuire a conservazione che non ad influsso toscano (che per la Corsica dell'epoca vuol dire essenzialmente pisano)" [If Pisa (the Pisan variety), until the 13th century, has changed to the innovative type, the presence of the order *lo+mi* in Corsican can probably be attributed to conservation and not to Tuscan influence (for Corsica at this period this would mean basically Pisan); KAN] (Loporcaro 2009: 122, fn.21).

(12)   *dàllumi*   (Rohlfs 1941: 21)
         give.IMP_it.ACC_1SG.DAT
         'give it to me'

(13)   un   la   ti   póssu   prummétte   (Rohlfs 1941: 21)
         not   it.ACC 2SG.DAT   can.1SG.PRES   promise.INF
         'I can't promise it to you'

The issue of the exact dating of the Latinisation and Romanisation of Corsica is widely discussed in the literature. According to Nesi (1992) the Romanisation and Latinisation of the island took place from the 3rd century BC onwards. Arrighi (2002) instead claims that Latinisation took place in the 4th and 5th centuries, explaining it as a result of migration of big crowds. In fact, the low number of persons originating from the Italian mainland, and the absence of documentation, gives rise to doubts about any Latin in use in Corsica in the first centuries AD. Corsica's Tuscanisation is attributed to

the Pisans' conquest of the island (starting in 1120 AD), a view underpinned by the structural similarities between Tuscan and Corsican ('Corsican being so similar to Italian'). What remains unclear about this hypothesis is how the island could have been Tuscanised within one century by a few (Tuscan) Italians settling on the island around 1000 AD (cf. Marchetti 2001: 11).

3.2.2   *Tuscan times: Corsica under Pisa's and Genoa's rule (1120–1768)*

In 1016, Pisa and Genoa liberated Sardinia and Corsica from the Moorish. By papal decree, Corsica came under ecclesiastical administration in 1120, belonging then to the diocese of Pisa. The Pisans influenced various aspects of the island's life including religion, trade and agriculture (cf. Bottiglioni 1957b: 159).

As Corsica was a strategic point in the Mediterranean Sea, Genoa and Pisa entered into war. In the course of the Genoese–Pisan Wars, Pisa lost the territory of Corsica after the Battle of Meloria in August 1284 and consequently the whole island became Genoese territory. But before that, offensives by Ligurians had already taken place; in 1195 Bunifaziu was conquered by Ligurians.[39] When the Ligurians ruled over the island, existing towns expanded and new ones – Calvi, Bastia and Aiacciu – were founded along the coastline, and populated by Ligurians coming from the mainland. At first, Corsicans were not allowed to settle in these towns (cf. Arrighi 2002: 17).

The following two centuries were turbulent times in Corsica, influenced by invasions of various peoples. The towers built for defence, and ruins thereof, can still be seen along the coast. Soon Corsica became an integral part of the Genoese trade policy; generally, the Genoese period was marked by a prospering together of the Corsican and Genoese aristocrats, interrupted only by a first attempt by Sampieru Corsu[40] to liberate the island with the help of the French (in 1553) (cf. Ottaviani 2010: 31, 32). The period from 1567 (the last battle of Sampieru Corsu, at that time already without any allies) to 1730 (the first riots by the Corsican population against foreign rule) was marked by demographic changes as well as progression in agriculture and commerce (cf. Arrighi 2002: 18). During the Genoese rule Tuscan remained the official language (cf. Dalbera-Stefanaggi 2002: 3).

In 1729, the first riots arose among the Corsicans against the foreign rule because of excessive taxes and the incompetence of the Genoese in handling acts of revenge between the clans (cf. Arrighi 2002: 18). From 1730 onwards, for almost forty years, the Corsicans fought against Genoese rule and for their independence. After a short reign of King Theodore of Corsica (1936–1743), Pasquale Paoli, a Corsican patriot and the subsequent head of government, took over the leadership in the

---

[39]   In the old centre of Bunifaziu, a Ligurian based variety is still spoken today (cf. Dalbera-Stefanaggi 1997: 303; Dalbera-Stefanaggi 2002: 116–120).

[40]   Sampieru de Bastelica, known as also Sampieru Corsu, (1489–1567), was a Corsican *condottiero* 'warlord', and is nowadays considered as one of Corsica's earliest leading figures in the battle for national independence.

struggle for freedom. His efforts resulted in the proclamation of Corsican sovereignty in November 1755. It lasted only for fourteen years (1755–1769), until the final battle of Ponte Nuovu in 1769 against the French. Despite the period of independence being so short, Paoli achieved preliminary changes in the economy and education: the construction of a Corsican fleet and port in L'Île-Rousse and the inauguration of the university in Corti opened in 1765 (and closed down when Paoli was exiled in 1769). During the revolution and the leadership of Pasquale Paoli, the Corsican-Italian diglossia was not under discussion:

> Un des paradoxes de la littérature insulaire consiste dans le fait que la Corse ait pu s'organiser en nation indépendante, sans qu'il y ait jamais eu de réflexion sur la langue. La période de la Révolution corse, qui va de 1729 à 1769, a donné lieu a un certain nombre d'œuvres typiques sans rien modifier de la structure de la littérature traditionnelle. (Piazza 2009: 187)

> [One of the paradoxes of the island's literature consists in the fact that Corsica was able to organise itself as an independent nation without ever having to think about the language. The period of the Corsican revolution, between 1729 and 1769, gave rise to a number of typical works without changing the structure of the traditional literature; KAN]

The issue of an independent Corsican language did not arise amid the eventful circumstances of the independency. This is interesting given the importance the language issue had for the separatist movement in the 20th century.

 During the 17th and 18th centuries, education provided on the island was scarce and concentrated in Bastia.[41] For higher education Corsicans went to Italy, nearly exclusively to Pisa. Writers developed close relationships with the Italian mainland (above all Genoa, Venice and Rome) as a result of their education, or long-term journeys, which often led to the permanent residency of Corsicans in Italy. The cultural-educational orientation towards Italy had not only linguistic reasons, but also considerably practical ones. Corsica's external contacts depended exclusively on maritime relations; in winter the island was cut off from the French mainland for months due to unsteady weather conditions, so the comparatively short distance between Corsica and the Italian mainland promoted Italy as the main location for higher education.

 During the domination by Pisa (1077–1284) and Genoa (1284–1768), neither tried to impose

---

[41] For the 17th century there were only two institutions of higher education, the *collèges*, lead by Jesuits and exclusively frequented by the urban patricians.

their language on the island. Nevertheless, Nesi (1992) regards this period, in particular the Pisan rule, as linguistically the most important, especially for the northern Corsican varieties.

Under Genoese dominion the linguistic system of reference did not change much. Trade with Tuscany, already developed under Pisan rule, continued, and Genoa itself used the Tuscan-based written language as lingua franca in its trade relations (cf. Nesi 1998: 113). Because of the systematic occupation of the island from north to south, the north being geographically closer to the Ligurian mainland linguistically, the northern Corsican substrate is more influenced by the incoming Ligurians and their Tuscan.

For centuries, the linguistic situation on the island was characterised by a stable diglossia: Tuscan Italian, the *lingua dei signori*, was considered the language of literature, education and administration, i.e. the written domains or domains of communicative distance (*Distanzbereiche,* cf. Koch & Oesterreicher 1985), whereas Corsican was used in all other contexts of everyday life, i.e. the spoken communication domain (*nähesprachlicher Bereich*).[42] Nevertheless the Corsican–Italian diglossia was only practised by a relatively small number of people. The public at large, not being alphabetised, was monolingual Corsican (cf. Dalbera-Stefanaggi 2002: 15). In a general sense, the diglossic situation was admitted as normal; Corsican and Tuscan were not considered as distinct languages, but rather as different levels of the same idiom (cf. Thiers 2008 [1989]). Therefore, the need to elaborate a written Corsican did not arise. Also, the high inter-comprehensibility of the two varieties may have played an important role in keeping the situation stable.

Considering the pre-French situation of Corsican and Italian in detail, the two language variety match one domain: Corsican was used in the spoken domains, Italian in the written domains, with only smaller exceptions. According to the four-stages typology, elaborated by Marchetti (1989), Italian as a spoken medium was only used in church, where it was complemented by Latin for the liturgy (cf. Nesi 1994; Blackwood 2004a: 135).

3.2.3 *The French period (I): from 1768 until the end of the 20th century*

The French 'linguistic conquest' can be said to have passed through different domains as time went on. It began in the administrative domain where it was directed against (Tuscan) Italian, and moved on to the educational domain. Only when compulsory schooling was established (1882) did it turn against Corsican, by infiltrating private domains such as family life.

When Corsica became part of the French territories in 1768 with the Treaty of Versailles, France "followed a non-interventionist linguistic policy, satisfied that Corsica was a part of France

---

42   Diglossic situations with a spoken local variety and written Tuscan Italian were also common on the Italian mainland until the middle of the 19th century.

as long as those who wielded power spoke French" (Blackwood 2004a: 136). During the following two decades, there were no significant changes in the given Corsican–Italian diglossia. The French language arrived late on the island, in comparison to the political affiliation to France. Only with the French Revolution, and the Jacobin policy of 'French only' after 1792,[43] did the linguistic situation on Corsica change significantly. The linguistic 'unification' of all inhabitants of the French territories by a common language was considered a threat for political stability (cf. Fabellini 2010: 49, 50; see also below). In the first decades after the French Revolution the language policies imposed by France were not directed against Corsican, but aimed to suppress Italian (cf. Blackwood 2004a: 133).

At the beginning of France's sovereignty all official documents were published in French and Italian. In 1790 documents were still bilingual, e.g. the Corsican Civil Code. The use of Italian was only interdicted for any official documents in 1852. It had no direct impact on the use of Corsican, since such documents never had been published in Corsican (cf. Blackwood 2004a: 140).

Italian remained the language of higher education until the first decades of the 19th century. In 1826, Italian was prohibited as language of instruction in colleges and only French-speaking teachers were nominated (cf. Marchetti 1989: 112, 113). Yet, the overall number of schools was limited and school attendance of children in rural Corsica, where the majority of the population settled, was modest due to the still-deficient road network and the need for children's labour on family farms. Formal education played a small role in the conquest of language domains in the countryside until the late 19th century when school became free, compulsory and secular (in 1882, according to the *Loi Ferry*) (cf. Blackwood 2004a: 143).

Only with the French education policy was a geographically decentralised primary school system established in Corsica. As languages of instruction, neither Italian nor Corsican have ever been of importance in basic schooling. French-speaking teachers were sent to the villages, where they found a nearly monolinguistic Corsican-speaking population; children who had grown up monolingual in Corsican when entering in school were suddenly confronted with a (mostly also monolingual) French-speaking teacher. In and out of class, children had to expect punishment for speaking Corsican. Furthermore, official reports from teachers testify to the use of Corsican even in previously Tuscan-speaking domains, such as debates of the parish council: "This use of Corsican could be understood in terms of its speakers using Corsican to fill […] domains previously occupied

---

[43] For contemporaneous documents about the Jacobin language policy, see Bertrand Barère de Vieuzac (1794) *Rapport di Comité de salut public sur les idiomes*, and Henri 'Abbé' Grégoire (1794) *Sur la nécessité et les moyens d'anéantir les patois et d'universaliser l'usage de la langue française.*
For more about the French language policy in general, see Ager (1999).

by Tuscan Italian" (Blackwood 2004a: 145).

The stigmatisation of Corsican (by the French) in the context of education was driven by two rationales prevalent at the time: the fear of counter-revolution,[44] and the assumption that language acquisition by children is negatively influenced by learning more than one language simultaneously. These rationales served as official excuses to banish Corsican from primary schools. In fact, then and now, children's education is one of the easiest and most effective ways to induce language shift (cf. Blackwood 2004a: 146). It ensures linguistic, but also cultural, influence on a whole generation at a very young age and through them, also on their families.

In spite of the fact that Italian lost domains of influence, "Italy remained a point of reference for Corsicans [until the occupation by the Italian fascists in November 1942]" (Blackwood 2004a: 143), in particular in the domain of higher education. In 1829, Corsicans amounted to 25% of all students at the university of Pisa (cf. Fusina 1994: 33). France had no power to change the principle causes of this intense exchange between Corsica and Italy, i.e. Italy's geographic position, and Italian and Corsican belonging both to the group of Italo-Romance varieties with the resulting ease of learning and using Italian.

Employment by public institutions provided and organised by the French state offered another opportunity to influence the linguistic behaviour of Corsican society beyond the reach of educational programs or administrative regulations (cf. Blackwood 2004a: 146). Most important in this regard were the civil service and public administration. These held the prospect of financial security, for a society which was mostly dependent on agricultural labour with all its uncertainties. With these new working opportunities, French gained importance for the rural population where Italian had never been of great importance. The economic opportunities offered by employment by the French state caused a non-forced extensive language shift.

Different social groups changed their language behaviour according to different preconditions. The urban intelligentsia and professional classes abandoned Italian as a medium of cultural production in favour of French; among them, Corsican was always limited to folk literature, poetry and music; other forms of written and spoken culture were undertaken in Italian: "Therefore, the switch to French served only to decrease the number of functions for which Italian, and not Corsican, was used" (Blackwood 2004a: 139). The urban middle class used French in particular to profit from

---

[44] France was suffering from fear of a counter-revolution (cf. Marcellesi & Thiers 1988: 818; Blackwood 2004a:140), not only in Corsica but in all its new conquered territories. Two conflicting ideas were behind this fear: first, Corsican could be used by Corsican nationalists to unify the population by bringing into relief the differences between Corsica, the Corsicans and France and the French; second, Corsican was the only language most Corsicans could understand and therefore be reached with. For more about insecurity due to language practices and the resulting measures taken by the French government, see Ager (1999: 8–14).

the economic advantages linked to employment by the French state, as did the rural population; they also abandoned Italian as a cultural language in favour of French. In addition, families often gave up Corsican: "[They] began to pass on French first and foremost to their children" (Blackwood 2004a: 146). For more than a century, the traditional domains of Corsican were scarcely affected by the imposed French policy. The only domain in which French acted upon Corsican was that of employment.

Language use in church and for religious ceremonies in general remained almost unchanged until the beginning of the 20th century. Although preachers had to pass a French language exam, Italian remained the language in use in non-formal (the liturgy was held in Latin) contexts. This allowed for better communication between preacher and community, in particular in remote rural areas. But also, Italian-written calendars and hymns could be found at the turn of the 20th century (the last catechism written in Italian was printed in 1912 and sold until the late 1920s). The final banishment of Italian in the domain of church and religion took place as late as in the 1930s, when Corsica turned away, politically and culturally, from Italy because of the Italian fascist conquest policy (cf. Casta 2003: 137, 138).[45]

With regard to the aforementioned diglossia, the situation in Corsica changed with the arrival of the French. At first, Corsican remained dominant in the spoken domains, while Italian and French shared the written domains with changing dominance relations (see Table 10, below). Slowly, French gathered influence but had no proper domains of its own during the first few decades; even official documents were still published with Italian translations.

The dominance relations within the diglossia can be illustrated in terms of the complex mechanism of *Überdachung* (in the sense of Kloss 1967). During the decline of Italian, until its complete suppression, the homo-genetic *Dachsprache* of Corsican (i.e. Italian) was replaced by a hetero-genetic one (i.e. French). In general, *Überdachung* refers to the situation in which a standard language (e.g. Italian) 'roofs' one or more non-standard varieties (e.g. Sicilian, Neapolitan, etc., and, formerly, also Corsican). The *Dachsprache* (e.g. Italian) may serve as the language of written domains, while the non-standard varieties are dominant in the spoken domains. Thus, the

---

[45] Citation: "La grande cassure linguistique que sein de l'Église de Corse s'est opérée vers 1926. Les prêtres ont cessé de prêcher en langue vernaculaire. […] Les prêtres de cette époque que j'ai pu interroger ont tous eu la même réponse: 'les gens n'en voulaient plus l'italien'. […] Néanmoins, dans la plupart des paroisses rurales, les curés n'ont cessé de prêcher en langue locale jusqu'à la déclaration de guerre en 1939." [The great linguistic break within the Curch of Corsica took place around 1926. The priests stopped preaching in the vernacular. […] The priests I interviewd at that time all had the same answer : 'the people didn't want Italian anymore'. […] Nevertheless, in most rural parishes, the priests did not stop preaching in the local language until the declaration of war in 1939; KAN] (Casta 2003: 137, 138).

Dachsprache and the 'roofed' language variety/varieties are homo-genetic languages (e.g. belonging to the Italo-Romance language family). If a variety loses its homo-genetic 'roof', due to political changes for example, then the variety is considered a roofless, i.e. *dachlose*, variety (cf. Kloss 1976; 1978).

Table 10 illustrates the diglossia and its changes in Corsica until the point that the diglossia no longer existed. It shows the process of *Überdachung* by the French language, from the entry of French into the diglossic Corsican-Italian system until the complete supersession of Italian.

Table 10: Dominance relations in the diglossia (according to Fabellini 2010: 90; adaption and extension: KAN)

Legend:

- **A** or **a** (bold letters)   **Italian**
- A or a   **Corsican**
- B or b   **French**
- *   the two varieties to the left and to the right of the * stand in a diglossic situation
- | the varieties to the left and to the right of the | do not stand in a diglossic situation, but are both 'present' in Corsica at the moment specified in the timeline

Capitalisation/lower case: Gradient of dominance:

- capital letters (e.g. **A**, A, B)   dominating variety
- lower case letters (e.g. **a**, a, b)   dominated variety

Timeline	Until 1768 (arrival of the French):	Early French rule (until the first quarter of the 19th century):	First half of the 19th century:	Second half of the 19th century:	End of the 19th until beginning of the 20th century:	20th century, until today:
Diglossia	**A***A	**A***A | b	**a** | A*b	**a** | A*B	A*B	a | B
Description:	Italian (**A**) and Corsican (A) are in a diglossic situation; each variety dominates in its own domains (written or spoken).	Italian (**A**) and Corsican (A) are still in a diglossic situation; both languages dominate in their domains. French has no domain of dominance (b).	Italian (**a**) loses its dominance in the written domain; replaced officially by French (b). Consequently, French (b) enters in a diglossic situation with Corsican (A). Corsican still dominates the spoken domain.	Italian (**a**) is in continuous decline, without any proper domains. Corsican (A) still dominates the spoken domains. French (B) dominates the written domains.	Italian is supplanted in every official domain. Corsican (A) is still dominant in the spoken domains, and stands in a diglossic situation with French (B), which dominates the written domains.	Corsican (a) is in decline; no dominance in any domains. No diglossia with French (B), as French dominates almost every domain.

In general, the *Überdachung* of Corsican by French, a hetero-genetic language, is considered positive for the language-*Ausbau* of Corsican (cf. Muljačić 1989: 270; Thiers 1993: 254; Dalbera-Stefanaggi 2002: 4). For centuries, the Corsican–Italian diglossia was considered a '*muttersprachlich[es]*

Diasystem' [a native-language diasystem; KAN](Fabellini 2010: 88),[46] i.e. the pairing of spoken Corsican and written Italian was considered natural (see section 3.2.2 ; also Thiers 2008 [1989]) as a continuum, and the active *Ausbau* was not considered necessary (cf. Kailuweit 2014). Only as a result of the official interdiction of Italian and the implementation of French, i.e. the *Überdachung* of Corsican by French, did the will to develop and establish Corsican in the written domains arise.

The First World War, and the period shortly after it, had a big impact on Corsican. If we adopt the idea that male speakers, as opposed to female speakers, are associated with the maintenance of language traditions (cf. Blackwood 2004a: 147), it was a big loss for the Corsican-speaking community when a huge number of Corsican men enlisted for the French army (estimates vary between 20,000 and 50,000). About 12,000 Corsicans died in the war. Those returning home came back speaking French. Moreover, thousands of Corsicans emigrated to the more industrial centres in France, or even to other European and overseas countries, for economic reasons. In the 1920s, the first nationalist parties were founded, e.g. the *Partitu Corsu d'Azione*, which became the *Partitu Corsu Autonomistu* in 1927. It was organised by Petru Rocca and the *Muvristi*, who were named after the newly founded magazine *A Muvra* (cf. Pomponi et al. 2009: 68). The movement was suppressed shortly after, in 1939.[47]

In this way, beginning in the late 19th century, and reaching an initial peak between the two world wars, the two main issues relating to Corsican developed: linguistic activism, and lobbying for political separation (cf. Blackwood 2004a: 148).

In the 1930s, in line with Italian fascists' theory of irredentism (see Ettori 1979; Thiers 1993 for a discussion of the sociocultural implications), Italy tried to strengthen relations between Corsica and Italy. Corsican students were awarded grants for Italian universities, and Italians immigrated to Corsica. This can be considered Italy's early exertion of influence on the island. Relations between Italy and Corsica broke off with the occupation of Corsica by the Italians in November 1942.

---

[46] Language as a diasystem – in the sense of Weinreich (1954) – refers to a 'system of systems'; two or more languages within a multilinguistic community are linked together into a diasystem. By calling it a *muttersprachlich empfundenes Diasystem,* Fabellini (2010) refers to Corsican and Italian both belonging to the Italo-Romance language family and the high degree of (inter-)comprehensibility.

[47] Citation: "Ce désastre […] entraîne le développement d'un premier nationalisme, exprimé en particulier, par la revue *A Muvra*. Si ce courant n'a pas d'effet au plan électoral, son influence idéologique est importante. Mais son rapprochement, pas toujours conscient, avec le fascisme italien, ainsi que la pression d'un groupe plus radical de jeunes poursuivant leurs études en Italie, discrédite ce 'corsisme' et à travers lui, toute revendication identitaire jusque dans les années 50." [This disaster […] led to the development of a first nationalism, expressed in particular by the magazine A Muvra. If this current had no effect on the electoral level, its ideological influence was important. However, its rapprochement, not always conscious, with Italian fascism, as well as the pressure of a more radical group of young people pursuing their studies in Italy, discredited this 'Corsism' and through it, any claim to identity until the 1950s; KAN] (Arrighi 2002: 22).

Blackwood (2004a: 149) states: "When Italian troops invaded and annexed Corsica, the romantic notions of separated brothers united again evaporated as powerful overlords exploited the resources of the island, mistreated the Corsicans, and subjugated the island's citizens". Any language activism was gravely harmed by this separation: "Corsican, as an Italo-Romance language and one closely related to Tuscan Italian, was identified with the Occupation, the poor treatment of the islanders, and the misplaced loyalties of the 1930s, when bonds between the island and the peninsula had been strengthened" (Blackwood 2004a: 150).

The defence against the Italian invasion and the final liberation from the Italians with the help of the French, led to a certain attachment of Corsica to France. Again, the number of Corsican emigrants to the French mainland rose significantly.

Linguistic development in the second half of the 20th century is marked by active language planning. Since the 1950s, various legislative steps have been taken on a national and international level to establish and secure the current status of Corsican as an official regional language of France. Different institutions in Corsica have been involved in the process of language planning. The focus here was placed on education, as a principle direction of language planning in Corsica.

The most important law in this context is the *Loi Deixonne*, ratified in 1951. At that time, Corsican was not included among the recognised French regional/minority languages, i.e. Breton, Basque, Catalan and Occitan. Only when the law was modified in 1974 by *Directive 74-33*, was it officially declared a regional language of France, together with Alsatian and Flemish. On an unofficial level, evening classes and student discussion clubs had already been established since the late 1960s.

Besides language planning, demographic changes also characterised the 1960s in Corsica. As a consequence of the Algerian war (1954–1962), the French authorities settled (and resettled) many immigrants from northern Africa in Corsica, among them Algerians but also Corsicans (called the *colons*) deployed with the civil service in northern Africa. This resulted in remarkable demographic changes. The population grew from 172,899 inhabitants in 1962, to 215,291 inhabitants in 1975 (numbers from France's National Institute for Statistics and Economic Studies, INSEE [2006]; cf. Blackwood 2007: 22). Alongside the sheer number of newcomers, the treatment they received from the French government also caused resentment against them: they were given land, loans and subsidies to cultivate the east coast while established Corsicans had to manage without this help:

> This hostility to the colons and the immigrants from northern Africa […] fanned the
> flames of regionalism […] Resentment was brewing not only between elements of

Corsican society (above all, those in the regionalist movements) and the French State, but also between a section of the island's population and the Arab immigrants (Blackwood 2007: 23).

Regarding the Corsican language, the immigrants had no major interest in learning Corsican – a small minority language of limited use in public domains – as they already spoke French, the national language.

In the 1970s, language activism strengthened, first through the foundation of the *Scola Corsa*, an association based on language activities. This was the first official organisation working for the Corsican language, its maintenance and distribution. In the same period, *L'Association pour la Défense de l'Étude de la Langue Corse de l'Est et du Centre* (ADECEC) was also established. The legislative basis changed when Corsican was included among the official regional languages of France by the modification of the *Loi Deixonne* (originally from 1951) in 1972 (see above). This was followed by the *Loi Haby* (in 1975) that assured teaching of Corsican for the duration of obligatory public schooling. These two laws were significant for Corsican: for the first time, Corsican was officially included in a domain from which it had previously been excluded, and where it got proper support.

At the same time, an enforced nationalism arose with the foundation of the *Action Régionaliste Corse* (ARC) in 1975 in Aleria, followed by the *Fronte di Liberazione Naziunale Corsu* (FLNC) in 1976, which acted clandestinely. As Arrighi (2002) states: "Expression au départ d'un sentiment identitaire profond, ce courant politique a connu de graves dérives. Il reste divisé entre un autonomisme hostile à la violence et un indépendantisme dont une partie soutient l'action militaire clandestine" [Based on a strong feeling of identity, this political tendency has known heavy fragmentation. It remains divided between a demand for autonomy while refusing any violence and a demand for independency whereof a subgroup approves secret military actions; KAN] (Arrighi 2002: 23; see also fn. 30). In the political atmosphere of the 1970s, when political national forces rose up and controlled all domains of everyday life on the island, the issue of the Corsican language – the idea of the return to an historical linguistic past, to the *lingua di i pastori* – played an important role (see also section 3.1.3). The insufficiency of Corsican to deal with topics beyond the private context of family, friends and traditional activities revealed the requirement of active language elaboration (see also section 3.1.3).

The 1980s were marked by the implementation of a structured teaching system in public education. During this time, the dichotomy between everyday language use and its assigned value increased: "A dichotomy had developed and continued to widen between the use of Corsican on the

one hand and its rising value and penetration into domains from which it had long been excluded on the other" (Blackwood 2007: 25; following Arrighi 1993: 103). The opening of new domains for Corsican neither resulted in an increasing everyday use of Corsican nor in an increasing number of speakers itself. Therefore, the idea of *obligatory* language teaching arose (Blackwood 2007: 25; the lively debate, see Jaffe 2001: 284–290). Nevertheless, this question was regarded with caution by Corsican society:

> Whilst not suggesting that a majority of the island's population were opposed to Corsican as a means of communication, one local perspective – amongst others – was that Corsican's place was not within the school, symbol of the French Republic, but rather that speaking Corsican was an individual, private matter. (Blackwood 2007: 25)

The supporters of obligatory Corsican classes achieved only partial success: an introductory course of Corsican language and culture has been obligatory for all students at the university of Corti since its reopening in 1981. Nowadays, learning Corsican is included in the curriculum of public schooling on a volitional basis; parents who do not want the 'one to three hours' of Corsican teaching for their children have to officially withdraw them from these courses.

Under Mitterrand's government (1981–1995), Corsica was chosen as the pilot scheme of the decentralisation policy and a Regional Assembly was established.[48] This Corsican Assembly is responsible for the island's administration, including the organisation of Corsican language education. Even though the establishment of the Corsican Assembly was an important concession to self-administration of the island and self-management regarding the language issue, the implementation of Corsican in everyday life encountered numerous barriers. For example, the registration of Corsican Christian names in the French administration was almost impossible. This strict limitation of any official use or application of Corsican by the French legislation and administration was considered typical for the 1980s and 1990s.

In 1991, a unique Territorial Public Authority, the CTC, was set up, based on the island's Special Charter (law of 13th of May 1991), to which the Corsican Language Charter is linked. Besides its responsibilities in the fields of economy, environment, sports, youth and culture, the CTC is the regional public authority in charge of proposing a development plan for the Corsican language, for

---

[48] By 1969, President de Gaulle had already proposed major changes in administration and organisation of the regions towards a decentralisation; he lost this referendum. His successor, President Pompidou, made only minor attempts to reinforce (the status of) the regions. Only Mitterrand started a more deliberated decentralisation policy for the regions.

investigation and financing of corresponding actions.[49] Even with all the legislative changes at implementing Corsican in public domains (such as school, media, administration etc.) its use is still optional in almost every context.

In the 1990s European policy regarding regional and minority languages became more important: the European Community promoted the importance and protection of regional languages across Europe by drafting the *European Charter for Regional or Minority Languages*. The effect this European policy had in France was twofold: on the one hand, French was given a special position by the state; on the other hand, France had to preserve its linguistic diversity. The inclusion of French into the French Constitution with a singularised status (French Constitution, § 2: "La langue de la République est le français"; 25.06.1992) and the thereby-inspired *Loi Toubon* (in 1994; 'not to make any concessions to the regional languages') could be seen as a step backwards (cf. Blackwood 2007: 27). But the *Loi Toubon* was rejected all over the country, and the changes and progress of decades were not, and could no longer be, stopped or cancelled.

Despite its defensive attitude, France was invited to join the *European Charter for Regional or Minority Languages*, but did not sign it until 1999. For France, the main problem with accepting and signing was the wording of the Charter; it acknowledged "the existence of groups of peoples within the state" (Blackwood 2007: 29). On the basis of this text passage, the Regional Assembly of Corsica requested the recognition of 'the Corsican people' in Paris, but the request was rejected by the French Constitutional Council, which ruled that there could be only *one* people in the French Republic. The Charter was therefore considered unconstitutional (cf. Blackwood 2007: 29). Only in 1997, under Jospin, did a commission examine the Charter again and decide that it was in fact acceptable. As a consequence, it was signed in 1999; its ratification is still outstanding.

3.2.4 *The French period (II): the 21st century and the current state of Corsican*

Since the turn of the century, Corsican has occupied a privileged position among the regional languages of France; it can be considered the most promoted and best protected regional language (cf. Blackwood 2007: 29, 30).

Regarding its *Ausbau*, Corsican is present in the public domains of culture, media and education: spoken in the programs of regional radio-stations (among others, *Radio France bleu Corse*, *Radio Corse Frequenza Mora*);[50] written in different print media (newspapers, journals, comics, etc.); used at cultural events (theatre, concerts, readings, and folk dance with song); found

---

49 The CTC is part of the decentralisation process planned by the French government; it "is unique in European France: no other region can boast its level of authority and administrative power" (Blackwood 2014: 64).
50 A study on Corsican in the media, in particular, in radio-programs from the 1980s, has been conducted by Thiers (2008 [1989]: 211–234).

in bilingual street and place-name signs and bilingual public information (e.g. signs showing opening hours or instructions, for example, for separation of refuse in public spaces). The number of public events where Corsican is used and promoted is increasing ('language days', conferences including the conference publication in Corsican). In the context of public administration, the CTC publishes some documents bilingually (French–Corsican) and offers 'linguistic help' for enterprises and corporations needing translations ("*Site un impresa, una cullettività lucale, un associu, un particulare? Avete bisognu d'un aiutu linguisticu?*" [Are you a company, a local collective, an association, a freelancer? Do you need linguistic help?; KAN][51]. The CTC's website is written predominantly in French, except for the language department's homepage, which is almost exclusively written in Corsican.[52] In 2013, the CTC also initiated a project on children's participation in its activities, the *Assemblea di i Ziteddi*; the related website is bilingual, with Corsican content mostly presented first.[53] In general, Corsican is present on the Internet, with blogs, newspapers, apps, and bilingual websites of cultural institutions. Audio-visual media, such as movies, computer-games and interactive teaching materials, are promoted.

The most important device in language planning has been teaching in public education, especially on the levels of pre-primary to secondary school, but also in higher education (university) and to a certain degree in adult education. Remarkably, Corsican was never part of the educational system until its implementation in the 1980s, either as a subject or as a language of instruction.

Since 1982, Corsican has been embedded in a structured way into the public educational system (for a detailed description of the single steps and actions see Fusina 2000: 4; Fusina & Arrighi 2012: 6–9).[54] The public school system is exclusively responsible for the teaching of Corsican children – in the absence of other well-organised partners. Nowadays, teaching of Corsican is offered at any level of public education (from infant school to university) (cf. Fusina 2000: 5 ff.).

As well as regular language classes at school, incentives have been created for young people to learn Corsican – for example, the *André Fazi Prize* which is awarded to the two pupils obtaining the highest marks in Corsican in the Baccalaureate examination. Also, outside of Corsica, language education activities can be found where there is adequate demand, such as in Bouches-du-Rhône for instance, where pupils are prepared for the Baccalaureate examination in Corsican.

At the university of Corti, about 100 students study *Corsican Language and Culture* at the faculty of humanities. For all students at the university of Corti, an introductory course into Corsican

[51] https://www.isula.corsica/linguacorsa/Serviziu-di-traduzzione_r17.html [accessed 13.09.2019].
[52] https://www.isula.corsica/ [accessed 07.12.2021].
[53] http://www.ghjuventu.corsica/ [accessed 07.12.2021].
[54] The publications (Fusina 2000; Fusina & Arrighi 2012) are the first and second updated edition of the *Mercator Regional Dossiers Series* regarding the Corsican language in education in France.

language and culture is obligatory. A total of about 3,000 students learn Corsican (cf. Fusina & Arrighi 2012: 29).

Teaching of Corsican in non-academic adult education has been in decline since the 1980s, probably because of the organisation of public education at all levels of schooling. Moreover, the opportunities offered by new media and technologies may have facilitated the development of extra-curricular educational activities. Nevertheless, certain traditional methods of teaching can still be found at the level of adult education: associations, such as *Scola Corsa* at the *Casa di a Corsica,*[55] offer language courses to adults; or the regional branch of the national distance learning centre (*Centre National d'Enseignement à Distance,* CNED) in Aiacciu which organises correspondence courses in Corsican. Outside of Corsica, regional associations in Paris, Marseille or Nice offer evening language courses alongside the cultural program.

Even with all these projects, programs and initiatives, undertaken to create proper space for the Corsican language within society and everyday life, the consequences of these measures cannot yet be foreseen or evaluated (see also Blackwood 2007: 31). Young people are rarely heard speaking Corsican among themselves even though its standing has improved compared to preceeding decades.

In spite of all the difficulties relating to the implementation of Corsican as an integral part of Corsican life, Arrighi (2002) judges the general situation as a positive one, in so far as the discussions are more about its implementation and less about its right to exist:

> Cette dramatisation parisienne de la question linguistique corse contraste fortement avec la sérénité qui préside désormais à son traitement dans l'île, depuis une dizaine d'années au moins. Le corse y est universellement défini comme langue. […] Si discussion il y a à son sujet, elle porte sur les modalités concrètes de sa mise en œuvre et non sur son existence, elle est pédagogique et non plus politique (Arrighi 2002: 8).

> [This Parisian dramatisation of the Corsican language issue stands in strong contrast with the cheerfulness which has characterised its handling on the island for more than a dozen years. Corsican is defined as language there […] If there are any discussions about it, then they are about the specifics of its implementation and not about its existence; the discussion is pedagogical and not political anymore; KAN]

In summary, we can register a continuous and vigorous commitment to Corsican on all levels of

---

[55] http://www.maisondelacorse.fr/?page_id=782 [accessed 13.09.2019].

public life: from private persons, institutions and cultural events up to a political and legislative framework to promote and protect the language. Most of these activities are initiated and realised by Corsicans in Corsica. Beyond the borders of the island, Corsican is found only rarely; Corsica, as a topic in the broadest sense, most often concerns (traditional) music or events regarding culture, geography or politics. We may also find some written documentation such as books or internet pages.

Among the different forms of evidence for a vivid Corsican culture, linguistic studies play an important role. Socio-linguistic studies and lexical–phonological studies (with a diatopical focus) predominate.

Hence, the goal of this study on Corsican DOM is twofold: on the one hand, to enhance the value of Corsican as a topic of scientific studies by providing new empirical data that allow for new approaches, i.e. studies on morphosyntax; and on the other hand, by analysing the data against the theory on (Romance) DOM systems, i.e. different semantic and syntactic approaches, to classify Corsican DOM among the Romance DOM-systems in general.

4 Corsican DOM: State of the art

The beginning of Romance studies dates from the first half of the 19th century (cf. Diez 1836; Meyer-Lübke 1890). Corsican became a topic in linguistics shortly after, in the second half of the 19th century (cf. Mattei 1864; Falcucci 1875). From the start of this latter period, the phenomenon of Corsican DOM was documented. But it was about another hundred years before DOM was officially included among the specific characteristics of Corsican, and declared one of the seven '*indicateurs d'identité*' identified by Marcellesi (1983: 399). Since then DOM has played a major role in linguistic studies, for example in the work of Thiers (1986), who surveyed the perception of diatopical variation on the basis of variety-specific and pan-Corsican phenomena, including DOM. Not until the late, when systematic language teaching emerged, has Corsican DOM also been documented from a prescriptive point of view. In short, many Corsicans and linguists working on Corsican are quite aware of the existence of the phenomenon of DOM, but much less aware of its distribution according to semantic, morphological and syntactic rules.

In the context of Romance languages, DOM is traditionally termed 'prepositional accusative', 'prepositional (direct) object', or 'accusative a'; the same holds true for Corsican linguistics. In reference books on Corsican, we find DOM listed among the prepositions, as one of the various functions of the preposition *à* – i.e. the marking of accusative, locative and dative objects (cf. Giacomo-Marcellesi 1997: 39; Damiani 1998: 27; Casta 2003: 114; Gaggioli 2012: 33; etc.). Sometimes DOM is also considered a syntactic idiosyncrasy in the context of the Corsican noun phrase (cf. Bottiglioni 1957b; Albertini 1972: 92; Giacomo-Marcellesi 1997: 13; Nesi 2002; Casta 2003: 119).

The following chapter on the state of the art concerning Corsican DOM shows how Corsican DOM is presented in the literature, looking primarily at nominal classes, as this is the criterion most often mentioned. The role of semantic and syntactic features and properties are also considered and data presented where provided. A more exhaustive list of examples from the literature is available in the appendix (Chapter 9). Thus, I start with the representation of Corsican DOM according to noun classes (section 4.1), followed by studies that take humanness into account, including some remarks on other semantic features (section 4.2), such as specificity and singularity. The third section (4.3) is dedicated to the other major issue in the context of Corsican DOM: the incompatibility of the *a*-marker and the definite article. These constitute the main topics relating to Corsican DOM. The subject of section 4.4 is Corsican DOM and verb semantics, followed by a look at theoretical accounts in section 4.5, in particular the disambiguation hypothesis put forward in order to explain Corsican DOM. In section 4.6, among the most current accounts on Corsican DOM, I consider

Giancarli's (2014) hypothesis, which combines aspects of transitivity and affectedness. Before I conclude the chapter with a summary in section 4.8, I give some information on prescriptive rules relating to Corsican DOM (section 4.7).

4.1 Which nominal classes display DOM?

From the earliest available documentation, the distribution of DOM in Corsican has been linked to the nominal class of the respective DO fairly explicitly, while the relevance of humanness has been demonstrated in a more indirect way, via, for example, the exemplified classes specified for humans (e.g. personal pronouns, certain lexical proforms, and proper nouns), represented by the existing language data.

In the following paragraphs, I will try to distinguish, on the basis of specific semantic and structural differences, between the noun classes relevant for DOM: proper nouns (including certain kinship terms); (full lexical) common nouns; and proforms.

Within the class of **proper nouns**, the literature most frequently mentions *a*-marking of proper nouns **which refer to humans** (cf. Bottiglioni 1933–1942; Rohlfs 1941; Damiani 1983; Marcellesi 1983; Thiers 1993; Chiorboli 1987, 1991a; Giacomo-Marcellesi 1997; Nesi 2002; Casta 2003; Gaggioli 2012; Albertini 1972), or to toponyms (cf. Rohlfs 1941; Damiani 1983; Casta 2003; Gaggioli 2012). Also, kinship terms are among the most frequently mentioned classes (cf. Rohlfs 1941; 1984; Thiers 1993; Casta 2003; Gaggioli 2012). Less often, authors mention names of animals (cf. Albertini 1972; Marcellesi 1986), or of religious figures, e.g. saints (cf. Bottiglioni 1933–1942), or other kinds of proper nouns such as names of months, weekdays or years (cf. Casta 2003). In the following I cite an example from Bottiglioni (1933–1942), showing the proper names of religious figures:

(14) *Prighemm' **a** gghiesu gristu e **a** **a** Madonna.* (Bottiglioni 1933–1942)

worship.1PL DOM Jesus Christ and DOM DEF.ART Madonna map no. 1904

'We worship Jesus Christ and the Madonna'

Bottiglioni states that the preposition appears before proper nouns and pronouns (cf. Bottiglioni 1957b [1932]: 121). No further details about the semantic features of the DO-referent, such as [human], have been given, even though all cited examples refer either to humans or to saints. Thus, on the one hand, example (14) demonstrates the expected *a*-marking pattern for proper nouns, *a*

gghiesu gristu.[56] But the second item in example (14), *a a Madonna*, shows an unexpected pattern: usually the simultaneous appearance of a definite article and *a*-marking is excluded; here we find the proper noun *Madonna* bearing a definite article *and* the DOM-marker *a* (for more about the incompatibility of DOM and the definite article, see section 4.3).

Giancarli's (2014) study is based on a database, for which he also provides some descriptive statistics on the distributional pattern of DOM within his corpus[57] (cf. Giancarli 2014: 7). According to this analysis, all proper nouns referring to humans and dates in Giancarli's corpus get *a*-marked in DO positions: "[…] tous les noms propres, toutes les dates, tous les possessifs (dans leur construction sans article) se construisent avec un AcP [*a*-marking]" (Giancarli 2014: 7). Giancarli explaings this as being due to the unique reference these *a*-marked DOs denote:

> Noms propres, dates en fonction argumentale d'objet et possessifs sans article ont pour particularité d'avoir pour référent un objet considéré comme unique. De même les termes pris comme noms propres: *babbu* (mon père)*, missiavu* (mon grand-père)*,* etc. (Giancarli 2014: 8)

> [Proper names, dates in object function and possessives without article have a common particularity: they have as referents objects which are considered unique. The same holds true for terms used as proper names: *babbu* 'my dad', *missiavu* 'my granddad', etc.; KAN]

A strong preference for *a*-marking is found with kinship terms that express a close (i.e. first or second) degree of kinship, i.e. mother, grandfather, sister, etc. If they are related to the participants of the conversation, they never bear an overt definite article to establish referentiality (as other common nouns require). These items are usually *a*-marked in DO-function; see example (15):

---

[56]   As stated in the Introduction, the spelling in the examples (here and throughout) correspond always to the original writing in the respective sources.

[57]   The study is based on a corpus of written prose texts, predominantly original Corsican texts (*A Travisagna* and *U Tavaru* by Renatu Coti; *Contra Salvatica* by Mathée Giacomo-Marcellesi, and *E fole di zia Anna-Maria* by Ghjuvan Ghjaseppiu and Anna-Maria Franchi) and one French text translated into Corsican (*Le petit prince – U principellu* by Antoine de Saint-Exupery, translated by Santu Casta).
In Giancarli's (2014) corpus *a*-marked proper nouns[human] amount to 47.1% of all tokens. Hence, they form the largest group of all *a*-marked DOs. Neuburger (2008) also confirms (likewise based on a written corpus) that *a*-marked proper nouns[human] are the most frequent noun class in written, natural language. The respective numbers can be found in Table (11), below.

(15)   *aghiu*   *vistu*   ***a***   *bàbbitu*   (Rohlfs 1941: 13)

       have.1SG   see.PAST.PTCP   DOM   father_POSS

       'I have seen your father'

Example (16) refers to a toponym, the city of Paris:

(16)   *cunoscu*   ***a***   *parigi*   (Rohlfs 1941:13, fn.1; source cited:
                                           Muvra 1931: 9; Rohlfs 1949b: 375)

       know.1SG   DOM Paris

       'I know Paris'

Even though Rohlfs does not explicitly refer to nominal classes in his description of Corsican DOM, but rather to the degree of animacy of the DO-referent, he illustrates the phenomenon with items of different noun classes. As well as those cited above, he also gives an example with personal pronouns (see example 20, below, with the proforms).

Casta (2003) has also focused on proper nouns with regard to DOM. Among the *a*-marked proper nouns, he not only includes human proper nouns, kinship terms, and toponyms, as in the examples above, but also years, names of months, and weekdays (*a*-marking "[d]evant les millésimes (sans article), les noms de mois ou de jours, traités en corse come des noms propres COD […]" Casta 2003: 120; see also examples in the appendix, Chapter 9). The examples and the phenomenon are part of his book on Corsican syntax. He does not refer to the (internal) syntax of the DP-/NP-DO.

Common nouns are rarely mentioned, in the literature about classes relevant to DOM, as, unlike proper nouns, they are systematically excluded from *a*-marking. The non-marking is explained with reference to syntax, and the incompatibility of the definite article and the *a*-marker (cf. Damiani 1983; Marcellesi 1983; Marcellesi 1986; Chiorboli 1991b; Gaggioli 2012; Giancarli 2014). Isolated examples are given to demonstrate rare exceptions, where the definite article and the *a*-marker appear simultaneously. This is observable in the following two examples, (17 and 18):

(17) *salutami*   ***ai***   *vostri*   *gucini*   (Rohlfs 1984)

       greet_1SG.CL   DOM_DEF.ART. POSS.2PL cousins

       'Say hi to your cousins from me'

(18) *Mintulete*   ***à***   *u*   *prufissore Ch. Rostaing.*   (Chiorboli 1991b)

       mention_2pl   DOM   DEF.ART.M.SG   Professor Ch. Rostaing

       'Mention Professor Ch. Rostaing'

Again, even for common nouns which are exceptions to the syntactic 'rule', the authors cite examples with human referents.

With regard to **proforms**, we find in the literature (examples with): *a*-marked personal pronouns (cf. Bottiglioni 1933–1942; Rohlfs 1941; 1984; Damiani 1983; Thiers 1993; Gaggioli 2012); interrogatives (cf. Bottiglioni 1933–1942; Marcellesi 1986); indefinites, such as *qualchidunu, qualchissia* 'someone, somebody' (cf. Gilliéron & Edmont 1914; Bottiglioni 1933–1942; Nesi 2002; Gaggioli 2012); universal quantifiers such as *tutti* 'all' (cf. Giacomo-Marcellesi 1997); and negative quantifiers such as *nimu* 'nobody' (cf. Giacomo-Marcellesi 1997).

     An *a*-marked personal pronoun can be found in Bottiglioni's (1933–1942) work, maps nos. 556 and 557 and Rohlfs (1941: 13). These are given in examples (19) and (20):

(19) *cunnoscu a*   *bbattistu ma un cerc'*   ***ad***   *ellu*   (Bottiglioni 1933–1942:

       know.1SG DOM Battistu   but not search.1SG DOM him   map no. 556/557)

       'I know Battistu, but I am not looking for him'

(20) *cercu*   ***a***   *boi*   (Rohlfs 1941: 13; 1949b: 375)

       search.1SG   DOM   you.2PL

       'I'm looking for you'

While most authors do not offer any explanation on the *a*-marking of personal pronouns, Damiani (1983) explains overt marking by emphasising the personal pronoun, hence explaining DOM as a morphological means for pragmatic processes: "le pronom est mis en évidence, il est prédédé de la préposition *à*" or "le pronom n'est pas mis en évidence […] il n'est pas précédé de la preposition *à*"

[The pronoun is highlighted, it is preceded by the preposition *à* […] the pronoun is not highlighted […] it is not preceded by the preposition *à*; KAN] (Damiani 1983: 18).

However, this claim disregards the fact that non-emphasised personal pronouns are usually implemented as clitics, which can be neither marked by *a* nor emphasised at all. Also, in other Romance languages with DOM-systems, only the full or strong form of personal pronouns are marked, and these are intrinsically correlated with emphasis or contrast (cf. Lyons 1999: 137; for Corsican: Romani 2005: 43).

Besides personal pronouns, interrogatives denoting human referents are also usually *a*-marked, as in example (21):

(21)   ***a***   *ccual'*   *agghiu*   *da siguidà?*   (Bottiglioni 1933–1942: map no. 1961)

       DOM   which   have.1SG   to   follow.INF

       'Which one do I have to follow?'

The earliest documentation of Corsican DOM is based on systematically collected data concerns indefinite proforms and is provided by Gilliéron & Edmont's (1914) ALF C. DOM is presented in two maps, nos. 376 and 799, exemplifying an indefinite proform: *conduire quelqu'un* 'to drive someone' (Gilliéron & Edmont 1914: map no. 376) and *haïr quelqu'un* 'to hate someone' (Gilliéron & Edmont 1914: map no. 799). On these maps the diatopical distribution of DOM appears scattered all over the island, and the single varieties showing DOM differ to some extent between maps nos. 376 and 799. This has been also noted by Bottiglioni (1957c): "[…] in esse appare soltanto qua e là in dodici punti della prima carta e in due della seconda, come fenomeno sporadico" [It appears only here and there in twelve places on the first map and in two on the second, as a sporadic phenomenon; KAN] (Bottiglioni 1957c: 122, fn. 10). Besides ALF C, only Bottiglioni's (1933–1942) ALEIC offers such a detailed description of the diatopical distribution of DOM (and other phenomena) on the basis of linguistic maps.[58]

One indefinite proform that is lexically specified for human referents, is *qualchissia* 'someone/somebody', which we can see in example (22), taken from the ALEIC, map no. 361:

---

[58]   The data of the NALC-project may contain data diatopically even more fine-grained and maybe includes also DOM-examples, but the data are neither annotated from this point of view, nor openly accessible (see also Chapters 3 section 3.1.5, and Chapter 5 section 5.3).

(22) [...] *o   ghe   gi   mandi   **a**   ccalchissia?*   (Bottiglioni 1933–1942:

     or   that   there send.2SG DOM somebody   map no. 361)

     'Would you like me to go, or are you sending somebody?'

Besides the proforms mentioned above, which are all specific human-referring lexical items, the universal quantifier *tutti* 'all' can be used irrespective of the degree of animacy. In the example given by Giacomo-Marcellesi (1997) it refers to humans; see example (23):

(23) *Mi   salutaret'**à**   tutti!*   (Giacomo-Marcellesi 1997)

     1REFL greet.2PL_DOM   all

     'Say 'hi' to everybody from me'

For negative quantifiers too, there are examples in the literature. These include *nimu* and *nisunu*, both meaning 'nobody':

(24) *un   vigu   à   nimu*   (Marcellesi 1986: 138)

     not see.1SG DOM nobody

     'I see nobody'

To summarise, animacy, or rather humanness, is even more relevant to proforms than it is for proper nouns as evidenced by the absence of any examples of proforms referring to animates or inanimates, even to demonstrate the impossibility of *a*-marking [animate] or [inanimate] proforms. All proform-examples without exception refer to [human] referents.

As we can see, then, none of the existing studies offers a complete picture of the distribution of Corsican DOM with regard to noun classes. The data cited do not cover all classes of nouns possible in DO-function. For a long time, the most complete picture has been offered by Marcellesi (1986); whose picture of Corsican DOM inspired the work of Neuburger (2008), and Neuburger and Stark (2014).

The data from Neuburger 2008, and Neuburger and Stark (2014), are part of a written corpus of genuine, non-translated Corsican texts. This collection includes novels, newspaper articles, and a conference paper, selected to cover different text genres. The studies categorised DOs according to their degree of animacy (human, animate, inanimate) and may be divided into the following nominal

classes: strong personal pronouns, proper names, definite pronouns, definite nominals (i.e. definite article/demonstrative/universal quantifiers followed by common nouns or kinship terms), indefinite pronouns, indefinite nominals (i.e. indefinite article plus common noun), quantified nominals, and bare common nouns (cf. Neuburger & Stark 2014: 376). The distribution of *a*-marking according to these authors is given in Table 11:

Table 11: Overview of the parameters investigated in the corpus study (Neuburger & Stark 2014: 376)

Legend
+	all examples of this class attested in the corpus are marked
±	marked and unmarked examples of this class are found in the corpus
–	no examples of this class are marked in the corpus
ø	no examples of this class are attested in the corpus

	Strong personal pronoun[i]	Proper nouns	Definite pronouns		Definite nominals	Indefinite pronouns		Indefinite nominals	Quantified nominals	Bare nouns
		anthroponyms, toponyms, kinship term[ii]	∀	demonstrative	art+N, art+poss+N, dem+N, ∀+N	negative pronoun[iii]	others	indef. art+N	numeral/*some/many* +N	
[+human]	+	+	+	–	–	+	–	–	–	–
	3(3)[iv]	39(39)	5(5)	0(7)	0(103)	4(4)	0(5)	0(17)	0(4)	1(16)
[+animate]	+	ø	ø		–	ø	–	–	ø	–
	2(2)	0(0)	0(0)		0(13)	0(0)	0(1)	0(3)	0(0)	0(5)
[+inanimate]	±	–	–		–	–	–	–	–	–
	11(13)	0(1)	0(9)		0(488)	0(4)	0(0)	0(13)	0(4)	0(53)

i Strong personal pronouns are only used to refer to human beings and therefore there are no attested examples in the classes of animate or inanimate referents.
ii Kinship terms with a definite article are attributed to the class of definite nominals.
iii *nisunu, nimu* 'nobody', *nunda* 'nothing'
iv The number before the brackets is the number of *à*-marked examples, the number in the brackets the total of examples in the sample. I.e. reading 'three examples out of a total sample of three are marked'

Table 11 shows that, proceeding from left to right, personal pronouns, like human proper nouns (even if used metonymically), always refer to human referents and are always *a*-marked,:

(25) *ellu*   *avia*   *lettu*   ***à***   *Fenimore Cooper*

He   have.3SG.IMPERF read.PAST.PART DOM   Fenimore Cooper

'He had read Fenimore Cooper'   (Neuburger 2008)

Names of animals, too, are consistently marked. Within the class of toponyms, out of a total of 13 toponyms, two appear without *a*-marking, whereas one is used in a metonymic way:

(26) *Enrico* *Marcia:* *hè*   *statu*   *u*   *primu* *à* *cantá*

Enrico Marcia be.3SG.PRES be.PAST.PART DEF.ART first   to sing.INF

à   *Solenzara*

DOM   Solenzara

'Enrico Marcia: he has been the first to sing [the song]   (Neuburger 2008)
Solenzara [location on the south-east coast of Corsica]'

The variation among kinship terms, noted by Marcellesi (1986) and described by Giancarli (2014) – i.e. they are commonly, but not always, used without a definite article – are also visible in this corpus. If used without a definite article, kinship terms are consistently marked, but they may appear with the definite article:

(27) *Tutti quant'è no' simu,*

salutemo   *l'omu*   *d'azzione* *è* *ø* *u*   *babbu* *di a negritudine*

Greet.1PL.PRES DEF.ART_man of action and   DEF.ART father of the *Négritude*

'Together, we all salute the man of the action, and the father of the *Négritude*' (Neuburger 2008)

The appearance of the definite article, instead of the *a*-marker might be explicable in terms of the metaphorical application of the term *babbu* 'dad' is applied in this context. In similar lines Giancarli's (2014: 7, 8) account follows along similar lines, re-iterating what has already been

said about kinship terms, above. In other words, in a given context, only those kinship terms that refer to humans related to the 1st or 2nd person (i.e. speaker or hearer), are used in a 'proper noun-like' manner, i.e. without a definite article. If they are used more 'common noun-like', i.e. if they are related to a third person (a third party), they bear a definite article (instead of the *a*-marker).

Within the class of proforms, certain DOs that refer to humans are *a*-marked, therse include universal quantifiers (*tutti* 'all') and negative pronouns *(nimu* 'nobody'), while pronominal demonstratives are never *a*-marked:

(28)   *Chjama tutti i principi di Grecia*

     per   scegli   ø   quellu   chì   serà u so ghjennaru

     to   choose.INF   DEM   who

     'He calls all the princes of Greece to choose the one who will be his son-in-law' (Neuburger 2008)

Of all noun classes, Marcellesi (1986) singles out demonstratives as the only 'class of transition', i.e. a class containing both *a*-marked and unmarked items referring to humans (for more detail, see section 4.2). In the corpus of Neuburger (2008), all pronominal demonstratives referring to humans are followed by a restrictive relative clause (cf. Neuburger & Stark 2014: 378). Whether this has an impact on the occurrence of the *a*-marking was not systematically investigated in this study. None of the nominals bearing an overt definite article or a demonstrative (also: plus possessive) are marked, regardless of the degree of animacy of the DO-referent: *Vigu ø l'omu* 'I see the man' (Neuburger & Stark 2014: 379; after Marcellesi 1986: 137).

Indefinite pronouns, such as *qualcosa* 'something', denoting inanimates, or *chalchissia* 'whoever', denoting humans, are not marked. Likewise, nominals bearing an indefinite article, are not *a*-marked. Adnominal numerals seem also to prevent the *a*-marking:

(29) *U Cicloppe li risponde ch'ellu si n'empippa di e so sciagure,*

    si   pesca   ø   due   omi   i sbatte di capu nant'à petra [...]

    ref.3SG   catch.3SG.PRES   two   men

    'The Ciclops responds to them that he does not care about their injuries, catches two men, beats them with their heads on a stone […]' (Neuburger 2008)

With regard to bare nouns, of the 16 bare nouns denoting humans, 15 are not *a*-marked. These 15 bare nouns consist of only two different lexemes: *moglie* and *maritu* 'wife' and 'husband', and appear in most cases in incorporation structures with the verbs *tene* 'to keep' or *piglià* 'to take'. Example (30) shows the exception:

(30) *Omu si dumandava cum'elli evianu fattu à truvà,*

    è   sopratuttu   à   tene   ø   moglie [...]

    and   above.all   to   keep.INF   Wife

    'One asked oneself how they managed to find, and above all to keep a wife […]' (Neuburger 2008)

4.2 What is the role of the feature 'humanness'?

The available data thus suggest that humanness is of importance to Corsican DOM, in that most examples of DOM refer to human beings. The animacy value [human] is usually recognised by the application of world knowledge, as with human proper nouns such as *Battistu*, or by the application of lexical knowledge, as with specified lexical items exclusively used for human referents, such as *nimu* 'nobody'.

However, some authors refer explicitly (or even exclusively) to animacy or humanness, often named 'personification'. Hence, DOM occurs with 'persons' and also objects, where a metonymic transfer and an interpretation as human are imaginable. This idea of transferred humanness is pursued by authors such as Rohlfs (1941) (for an account of transferred animacy, see also Yamamoto 1999, section 2.1). In general, Rohlfs is one author who does not refer to the nominal classes of the DO to explain the marking strategy, but motivates DOM with the animacy degree of the object referent: "[…] nella costruzione del complemento oggetto […] riferito a persona [the object] viene introdotto dalla preposizione *a* […]" [In the construction

115

of the object complement […] referring to a person is introduced by the preposition *a*; KAN] (Rohlfs 1941: 13). Nevertheless, in the examples cited by him, the respective DOs pertain to different nominal classes: the personal pronoun *boi* 'you[2pl]' and human proper names, kinship terms and toponyms (see examples 15, 16 and 20, above). In this way, the marking of toponyms is due to having "valore di nomi di persona anche" [value of personal names; KAN] (Rohlfs 1941: 13, fn.1).

Other authors follow a double line of reasoning, where personification or humanness is one of the relevant properties for *a*-marking. This is also Damiani's (1983) approach. His representation of DOM, in the chapter '*Le complément d'objet direct (C.O.D.) introduit par la préposition à*' (Damiani 1983), is the first detailed description of the phenomenon and is based on two lines of reasoning: semantics, i.e. the animacy of the DO-referent: "L'emploi de la préposition *à* comme signe de l'accusatif est une exigence grammaticale quand il s'agit de personnes" [The use of the preposition *à* as a sign of the accusative is a grammatical requirement when dealing with persons; KAN] (Damiani 1983: 17); and morphosyntax of the DO, i.e. the basic differentiation between nouns and pronouns (see also section 4.3).

Even though authors as Damiani (1983) recognise that Corsican DOM is multifactorial, i.e. by quoting animacy and syntax, the description remains rather vague. In these studies, humanness and syntactic structure do not seem to interact in a systematic way. Marcellesi (1986), with his first monothematic study on Corsican DOM, specifies the details in the following way: the distinction of [±human] of the DO-referent is neither crucial for proper nouns – those referring to humans are always marked, but toponyms or names of animals might also be – nor for personal pronouns. Pro- or enclitic personal pronouns are never marked, while full or strong personal pronouns always bear the *a*-marker. As full personal pronouns in Corsican do refer invariably to humans, and they are always *a*-marked, the *a*-marker does not assign or express humanness to/of the DO referent. According to Marcellesi, only within the class of (non-personal) pronouns does the *a*-marking indicates the distinction between a human and a non-human referent: "Désormais la règle est que l'on doit avoir *à* caduc [DOM *à*] quand le référé est humain" [From now on the rule is that one must have *à caduc* when the referent is human; KAN] (cf. Marcellesi 1986:133). In Corsican (and other Romance languages), most proforms have items lexically specified for humanness, e.g. negative quantifier *nimu* 'nobody' or indefinite *qualchissia* 'whoever'. Here too, then, *a*-marking does not provide the distinction between [human] and [inanimate]. Only a very few classes have exponents that are underspecified for animacy, e.g. the indefinite *unu* 'one' or the universal quantifier *tutti*, 'all',

already mentioned above.

Marcellesi (1986) is also one of the very few authors who mention any 'areas of transition' or domains where *a*-marking might be optional. First, the aforementioned demonstratives, pronominally used, with human referents; these might be used without *à*, for reasons of pragmatism, to express contempt: "[…] *quissu* non précédé de *à* caduc [DOM *à*] et référant en principe à un non humain pourra représenter une personne traitée de ce fait avec mépris" [The demonstrative *quissu* not preceded by *à caduc* and referring in principle to a non-human may represent a person treated with contempt; KAN] (Marcellesi 1986: 133). Thus, the use of demonstratives to refer to humans is already considered degrading. As for the second optional domain, Marcellesi mentions the interrogative pronoun *chi* that may occur without *à*, but he neither goes into detail nor gives any examples for these exceptions. According to descriptions in grammar books on Corsican (including also the therein cited examples), *chi* is considered the nominative interrogative pronoun, 'who/what', and may refer to humans or objects, while the accusative interrogative pronoun, *quale* 'who', is applied only with human referents.

Neuburger (2008), and Neuburger and Stark (2014), confirm Marcellesi's (1986) proposition regarding humanness in relation to noun classes: it is crucial in the domain of proforms. In general we advocate for a concept of individuation, based on Hopper and Thompson's (1980) approach that "a high degree of animacy (humanness) and/or referentiality contributes to the individuation of a DO (referent)" (Neuburger & Stark 2014: 383). This holds true for the representatives of the following categories: proper nouns, personal pronouns and kinship terms (without a definite article). These refer (mostly) to humans and are either inherently definite or/and refer in a definite–specific way in a given context.

Inspired by the cross-linguistic discussion on DOM (cf. Niculescu 1959; Bossong 1991; Lazard 1998; Aissen 2003), Giancarli (2014) investigates the importance of animacy (or humanness) to Corsican DOM and also takes into consideration definiteness, specificity and singularity (and the relevance of the overt definite article, see section 4.3).

Starting with animacy, Giancarli states that humanness of the DO-referent does favour overt marking; but out of a total of 189 DOs[human] only 55% appear with the *a*-marker, while

45% are without. As to definiteness, definite DOs[59] are also not consistently marked: out of 225 DOs[definite] 48.5% appear *a*-marked, while 51.5% are unmarked. Looking at humanness and definiteness together, 82.5% of all *a*-marked DOs in his corpus have a human and definite referent, whereas only 64.5% of all DOs[human, definite] are *a*-marked: "Ce qui ne fait donc pas du couplage animé humain + haut degré de détermination un bon critère" [This does not make the pairing 'human animated + high degree of definiteness' a good criterion; KAN] (Giancarli 2014: 5).

Specificity is investigated by Giancarli predominantly under the aspect of *genericity.* According to him, DOs with a generic interpretation never bear *à,* irrespective of their degree of animacy or the kind of article they bear, e.g. the definite article singular as in example (31) (but also plural articles or indefinite articles):

(31)   […] *si   pò   ancu   tumbà   ø   u   malatu*   (Giancarli 2014: 6)

one   can.3SG   also   kill.INF   DEF.ART.   patient

'[…] one can also kill the patient'

OR:

[…] *si   pò   ancu   tumbà   à   u   malatu*

one   can.3SG   also   kill.INF   DOM   DEF.ART.   patient

'[…] one can also kill the [specific] patient'

This generalisation is clearly reflected in Giancarli's corpus: "aucun example d'AcP n'est à interprétation générique, et tous les génériques s'emploient sans AcP" [No examples of AcP has a generic interpretation, and all generics are used without AcP; KAN] (Giancarli 2014: 6). By adding *à* to the DO *u malatu,* the generic reading of the aforementioned example might change into a specific reading, see option B in example (31): "[l]'ajout d'une préposition transformerait la référence en la rendant spécifique" [The addition of a preposition would transform the reference by making it specific; KAN] (Giancarli 2014: 6).

---

[59]   Giancarli does not offer a detailed either semantic or syntactic definition of what a definite DO is. Based on his comment, "[a]u sein de la référence spécifique en corse, le degré de detérmination est typiquement marqué par les déterminants […]" [With the specific reference in Corsican, the degree of determination is typically marked by the determiners; KAN] (Giancarli 2014: 4, fn.4), and the definition of "déterminant" (Giancarli 2014: 7) as definite article or a demonstrative, the definite DO may refer to nominals bearing an overt definite article or a prenominal demonstrative.

To summarise: the relevance of humanness is well established in the literature, even though it is often not addressed properly; most *a*-marked examples cited in the literature refer to humans, but this is not always made explicit. In addition, different properties interact for Corsican DOM, but not all features and properties are considered at all levels of description, with the result that the representation is often sketchy.

4.3 **The incompatibility of Corsican DOM and the definite article**

Bottiglioni (1957c) first noted the peculiar incompatibility of Corsican DOM with the definite article. In its basic conception, this incompatibility is usually linked to common nouns; of the noun classes mentioned above, this is the class most commonly associated with the presence of an overt definite article. By contrast, no information about bare nouns and their relation to DOM is present in the literature, with the exception of Neuburger (2008), and Neuburger and Stark (2014) (see also section 4.1). Common nouns are generally considered to be categorically excluded from *a*-marking due to the presence of the definite article. Other prenominal items, such as demonstratives, quantifiers, numerals or indefinites, are rarely mentioned (except by Marcellesi 1986; see appendix, Chapter 9, for an exhaustive list of his examples). Other DOs affected by DOM may include proper nouns, in particular toponyms, as well as names of animals or nicknames, bearing an overt definite article.

Hence, most authors observe that the overt definite article and the *a*-marker are incompatible, but do not scrutinise the function(s) and process(es) behind this incompatibility. One example of this is Damiani (1983), who makes the distinction – for the syntactic description of DOs – between those nominals which bear a definite article and those which do not. To illustrate the former, he provides the following example (32):

(32) *fighjula   i comunisti/   u jacaru/   u Tavignanu*   Damiani (1983)

see.2SG_there   the communists   the dog   the Tavignanu

'See there, the communists/the dog/the [river] Tavignanu'

Although Damiani does not refer to the semantics of the respective DO-referents, the different DOs show different values of animacy – i.e. [human], [animate] and [inanimate] – and represent different nominal categories – i.e. two common nouns, and one toponym. What Damiani considers crucial in this context, is the presence of the definite article (irrespective of

the semantic nature of the DO-referent):

> L'article exclut la préposition […] l'article et la préposition *à* dans les constructions où ces élément précédent le c.o.d. [the DO], sont comme en rapport d'équivalence, en 'distribution complémentaire'. On a l'un o l'autre élément bien qu'ils soient de nature différente […] (Damiani 1983:17).

> [The article excludes the preposition […] in constructions where these elements precede the DO the article and the preposition *à* are as if in equivalence, in 'complimentary distribution'. We have one or the other element although they are of a different nature; KAN]

Durand (2003) states that many Corsican toponyms require the article, in general suppressed in the Italian and/or French varieties, such as *i Fulélli* 'Folelli', *a Vulpaghjóla* 'Volpajola' . With regard to rivers, he states that river names do not bear an article: "*Gòlu, Tavignanu, Ristónica*" (cf. Durand 2003: 181). Despite Durand's (2003) work, the issue of definite articles and toponyms (or other geographically related proper names, e.g. rivers) remains rather vague. Damiani (1983) cites the river's name with a definite article (see example (32)) and without the *a*-marker. Giacomo-Marcellesi (1997) refers to the same river; here it does not bear a definite article, but is *a*-marked:

(33)   *U*   *focu*   *a*   *varcat'à*   *Tavignanu*

       DEF.ART   fire   have.3SG   overrun.PAST.PART_DOM   Tavignanu

       'The fire has overrun the Tavignanu'   (Giacomo-Marcellesi 1997: 42)

According to Marcellesi (1983, 1986), it is impossible to use proper names with definite

articles.[60] His study on so-called '*indicateurs d'identité',* or '*indicateurs linguistique de corsité*'[61] (Marcellesi 1983) includes seven linguistic characteristics of Corsican, of which the only morphosyntactic one is DOM. First here, and later in more detail Marcellesi 1986), he points out the difference between proper nouns that bear a definite article and those that do not. The former never get *a*-marked in DO-function, while the latter are consistently *a*-marked.

While other authors have established a relationship between the *a*-marker of DOM and the homophonous preposition *à* (used for locatives and datives), Marcellesi (1986) is the first who distinguishes the two items. As he observes, *à*$_{DOM}$ is different from the *à*$_{PREP}$ in that the former is sensitive to the (syntactic) structure of the following NE, i.e. the presence of a definite article, while the occurrence of the latter does not depend on this fact.

Marcellesi (1986) has focused on the morphosyntactic conditions for the (non)implementation of DOM, based on the distinction between proper nouns and common nouns (and humanness, as described in section 4.2). The simple classificatory distinction between proper noun (marked) and common noun (not marked) does not capture the pattern of Corsican DOM. Marcellesi cites proper nouns that include a definite article as an integral component, such as nicknames, toponyms (e.g. *U Conti,* a village's name), or animal names (e.g. *U Sciroccu,* the name of a horse) (cf. Marcellesi 1986: 137). These kinds of proper nouns are not *a*-marked if in DO-function. Nouns that refer, for example, to God or the saints get *a*-marked*,* but here too, the marking depends on the presence of an overt definite article. Kinship terms (in particular those referring to relatives of the first and second degree) are usually marked by *à,* as they usually do not bear a definite article and they have a specific reference.

Exceptions to the incompatibility of the *a*-marker and definite article are given by Chiorboli (1991b). He states that "l'article bloque d'ordinaire l'apparition de la préposition"

---

[60] Marcellesi's (1986) theoretical considerations have been confirmed by data from two surveys. The first is based on a written questionnaire taken from Rohlfs, which included 65 French sentences to be translated into Corsican. 22 Corsicans from different regions of the island participated at the survey. From the dataset, which included 1,212 sentences, only 45 sentences did not conform to the approach of Marcellesi. As the translation approach was considered problematic, on account of it being a written survey, Marcellesi conducted a second survey based on a 210-minutes recording of spontaneous speech in Aiacciu, with speakers from the region and from northern Corsica. These data did not reveal any divergence from the approach presented: "L'enregistrement de conversation spontanées à Ajaccio […] semble ne fournir aucune discordance avec le modèle proposé" [The recording of spontaneous conversations in Ajaccio […] seems to provide no inconsistency with the proposed model; KAN] (Marcellesi 1986: 135). The approach has also been tested on the basis of written texts included in the anthology of Corsican literature published by Mathieu Ceccaldi (1975). Marcellesi points out that only certain nominal classes in DO-function are represented in the dataset, both in the registrations and the anthology (cf. Marcellesi 1986: 135).

[61] Nesi, too, refers to the *indicateurs de corsité* with respect to DOM (cf. Nesi 2002: 963). But she also mentions, among the morphosyntactic peculiarities of Corsican, the particular word order of clitics, being accusative—dative, in pro- and enclisis, as opposed to the more common pattern of dative—accusative order.

[usually, the article blocks the appearance of the preposition] (Chiorboli 1991b: 77), but he also observes some variation, as in the following example (34), and example 18, above:

(34) *vurriami*   *dinò*   *ringrazià*   *à veru*

     would_like.3PL   besides   thank.INF   sincerely

     à   ***i***   *ripprisintanti*   *naziunalisti*   *asturiani*

     DOM   DEF.ART.   representatives   nationalist   Asturian

     'Besides, we would like to give sincere thanks to the representatives of the Asturian nationalists'

     (Chiorboli 1991b: 77) referring to Rotoly in *U Ribombu,* 25.03.1988

Regarding this example, Chiorboli comments: "on observe cependant dans l'usage des variation qui ne se laissent pas toujours facilement expliquer par le phénomène de la 'personification'" [However, there are variations in usage which cannot always be easily explained by the phenomenon of 'personification'; KAN] (Chiorboli 1991b: 77). This does not explain the exception from the aforementioned morphosyntactic rule.

     The data of Giancarli (2014: 7) also provides (very few) examples that contradict the rule of incompatibility: here, the DOs bear both the *a*-marker and the definite article. See the following example (35):

(35) *Ella,* ***à*** ***a*** *surella,* *ùn* *l'* *aiutava*   *mai.* (Giancarli 2014:7)

     she   **DOM** **DET** sister   not   her   help.3SG.IPFV   never

     'She never helped [her] sister'

Probably, the co-appearance of the two items in a written sentence does not correspond to its pronunciation: it may not perceptible due to haplology; no further explanation is given by Giancarli.

     Nevertheless, Giancarli (2014) takes up the complementary distribution between *a*-marking and the definite article, or rather the *déterminant*, which includes definite articles and demonstratives (see also fn. 17). Taking the rates presented by Giancarli (2014: 7) into account, there is a high degree of correlation between the presence/absence of a *déterminant* and the presence/absence of the DOM-marker: out of 119 *a*-marked DOs, 113 do not bear a *déterminant* (95%) and only six simultaneously bear the DOM-marker and a definite article

or a demonstrative: "[…] tous les objets sans AcP sont aussi munis d'un déterminant, et à partir d'une absence de déterminant on peut conclure qu'un AcP sera employé" [All objects without an AcP also have a determiner, and from the absence of a determiner it can be concluded that an AcP will be used; KAN] (Giancarli 2014: 7).

Neuburger (2008), and Neuburger and Stark (2014), conclude that the structural pattern, i.e. the incompatibility of the *a*-marker and the definite article, seems to be a "key property of Corsican DOM" (Neuburger & Stark 2014: 383):

> The Corsican object marker *à* appears to take as its complements only inherently definite and thereby highly individuated elements […] It does not take complex DPs, QPs or NumPs, where information about person […] and countability are marked in different functional heads and all amount 'compositionally' to a bigger or smaller degree to the [individuation] of the complex nominal (Neuburger & Stark 2014:384).

In summary, the authors of Corsican studies are quite aware of the structural idiosyncrasy of Corsican DOM that is the incompatibility of the *a*-marker and the definite article, although it receives less attention in the literature than noun classes or humanness. Aside from definite articles, other prenominal items are rarely taken into account.

4.4 Does verb semantics have an impact on Corsican DOM?

Giacomo-Marcellesi (1997) is the only one who mentions verb semantics as a relevant trigger for DOM. She considers DOM from two perspectives: she investigates on the one hand the semantic nature of the DO; on the other, the DO as the verbal complement. Regarding the latter, she states: "Quand il sont complément d'un verbe performatif ou d'un verbe de perception ou de mouvement, ces même noms sont introduits par la préposition tonique *à* […]" [When they are the complement of a performative verb or a verb of perception or movement, these same nouns are introduced by the tonic preposition *à*; KAN] (Giacomo-Marcellesi 1997: 13). The four examples she quotes include he verbs *chjamà* 'to call', *sènte* 'to listen', *vède* 'to see', and *varcà* 'to overrun'; the DOs in these examples represent the previously discussed classes of human proper nouns, kinship terms, toponyms, and a river name.

Giancarli (2014: 14, 15) also takes a look at verb semantics, and considers in particular verbs of perception. But, unlike Giacomo-Marcellesi (1997), he concludes that verb semantics has no effect on *a*-marking, as there are marked and unmarked examples in his corpus.

According to these authors, no distributional patterns of DOM can be correlated to verb semantics.

4.5 Does Corsican DOM disambiguate between syntactic roles?

Traditionally, the argumentation about disambiguation as a function of DOM is based on the correlation of the morphological similarity of nouns and the semantic similarities of the nominals' referents (for more detail, see section 2.2.2 on the disambiguation hypothesis). This is also the line followed by Rohlfs (1949a: 434–437; § 632; 1949b: 375). Disambiguation is the most traditional of all accounts on DOM in Corsican. To briefly recap, it refers to the presupposed need to overtly mark one of the two morphologically identical nominal arguments in a transitive construction, in order to disambiguate the two arguments from each other with respect to their syntactic function, i.e. subject and/or DO.

Chiorboli (1987) focuses on the morphological form of the nominal arguments from a diachronic point of view, and the flexible word order of the nominal in subject function. He claims that in modern Corsican the subject can be used either preverbally (subject-verb, or SV-order) or postverbally (verb-subject, or VS-order) (cf. Chiorboli 1987: 87). He cites the two examples *hà chjamatu à Petru* 'He/she has called Petru', versus *hà chiamatu Petru* 'Petru has called'. No further information is given on word order in general.

Giancarli too, mentions a somewhat flexible word order in Corsican: "[…] les arguments sont relativement mobiles puisque par rapport au verbe l'objet peut aisément être entéposé et le sujet postposé" [The arguments are relatively mobile since, in relation to the verb, the object can easily be put in the front or go after the subject; KAN] (Giancarli 2014: 10), resulting in object-verb-subject (OVS) order. Like Giacomo-Marcellesi (1997: 35–38) he does not refer to DOM.

Beyond flexible word order, Giancarli (2014) mentions contexts where disambiguation by overt marking would be very useful, but is not applied. This holds true if the subject and DO referent display similar semantic features, as a high (or low) degree of animacy and/or definiteness, the reasons traditionally invoked if arguing in favour of disambiguation (see also section 2.2.2). Nevertheless, according to Giancarli the hypothesis does not hold true for Corsican, as attested by examples from his corpus; in ambiguous contexts no disambiguation by *a*-marking takes place – see examples (36) and (37):

(36) *Infine, chì quand'i trè anni sò stati passati, ùn avaria cunnisciutu **ø** nimu **ø** a povera zitella di i quartieri i più povari d'Aiacciu […].*

   1. reading 'Finally, after three years had passed, nobody recognised the poor girl from the poorest neighbourhood of Aiacciu'

   2. reading '[…] the poor girl from the poorest neighbourhoods of Aiacciu did not recognise anybody'

(37) *Erani due sureddi ch'aiani spusatu **ø** due pastori.*

   1. reading 'There were two sisters who married two shepherds'

   2. reading 'There were two shepherds who married two sisters'

Giancarli (2014) mentions ditransitive constructions as another problematic issue for the disambiguation accounts, well known from accounts on other DOM systems (see also section 2.3 on syntactic accounts). Here, the unmarkedness of the subject and DO nominals is relevant. Theoretically, in ditransitive constructions, both the DO and the IO can be *a*-marked, resulting in morphologically identical structures. Their respective syntactic functions, as DO and IO, cannot be assigned properly: the *a*-marker may even cause ambiguity.

   According to Neuburger and Stark (2014), *a*-marking in Corsican is most persistent with proper nouns, universal quantifiers, and negative pronouns. The consistent marking seems exclusively driven by the DO's properties, and not by the subject properties, in the examples given (cf. Neuburger & Stark 2014: 383). Also, personal pronouns, consistently *a*-marked, do not confirm the disambiguation hypothesis, as their morphological forms differ for nominative and accusative, thus, no formal ambiguity occurs.

   As we have seen, disambiguation does not explain any of the Corsican DOM patterns in a straightforward way: there are contexts where the DO is *a*-marked, but disambiguation is already assured by other properties, e.g. the morphological distinction of personal pronouns; or where the *a*-marked DO appears in contexts in which there are other *a*-marked nominals, e.g. ditransitives (or even locatives); or where the DO is unmarked, and formally identical with the subject. In these situations, a formal disambiguation would be useful, but does not take place.

4.6 **Corsican DOM and Giancarli's *hypothèse de l'entonnoir***

Giancarli's (2014) approach to explaining Corsican DOM, called *hypothèse de l'entonnoir* ('Hypothesis of the Funnel') (Giancarli 2014: 11), combines aspects of the transitivity

hypothesis – the importance of the relation between subject, verb and D – and of Næss's (2004) affectedness approach (for more details about the theories, see section 2.2.3). Giancarli assumes that *à* diminishes the degree of affectedness of the DO. Thus, he inverses Næss's approach, where the most affected DOs get marked (cf. Næss 2004: 1203; see also section 2.2.3). But in fact, the examples given by Giancarli can be better explained by the existing approaches, be it in semantic or morphosyntactic terms.

4.7 Prescriptive rules: how to apply Corsican DOM?

All the studies I have so fas mentiones regarding Corsican DOM are explicitly or implicitly *descriptive*, with no *prescriptive* aspiration.[62] From a diachronic viewpoint, this is not surprising: there are only very few written documents in Corsican, much less any prescriptive grammar or language books. Also, from a synchronic point of view, the descriptive nature of linguistic information is not surprising. The *polinomia* concept promotes a polynomic approach to language learning and teaching which would contradict unifying, normative practices (see also section 3.1.2). Nevertheless, rules referring to the 'correct use of DOM' can be found in textbooks developed for language teaching at school or for independent language study.

One of the first prescriptive expositions of DOM was presented by Jean Albertini, author of several textbooks, conversation and writing guides, and dictionaries (French/Corsican), in his *Précis de grammaire corse* (Albertini 1972). Among the particulars of structure, he presents the various areas in which the preposition *à* is applied. Unlike French, *à* is applied with proper nouns in DO-function, but excluded in the context of common nouns or non-personified nouns such as unnamed animals ("animal non personnifié", Albertini 1972: 92). According to Albertini's description, the examples cited for DOM include only human proper nouns (*fieddiu a Angula* 'I see Angula', *approvu a Petru* 'I examine Petru', Albertini 1972: 92).

In the textbook *Le Corse sans peine*, a language course for independent language learning, DOM is barely mentioned: "Le complément direct avec *à*: […] il faut que le complément direct soit un nom propre […]" [The direct complement with *à*: […] the direct complement must be a proper noun; KAN] (Marchetti 1974: 173). The grammar book by Romani (2005), conceived as a reference book in school, does not address the topic at all.

---

[62] Or even excluding any prescriptive claim: e.g. in the preface of Casta's book, "Santu Casta s'est voulu plus déscriptive et analytique que prédictif et doctrinal" [Santu Sasta was more descriptive and analytical than predictive and doctrinal; KAN] (Casta 2003: Préface).

4.8 **Summary and open questions**

In summary, the state of the art on DOM-research in Corsican shows that DOM is a well-known phenomenon among linguists concerned with Corsican. Only a few studies have focused on DOM alone, but in general DOM is broadly attested in the linguistic literature.

Until the study by Damiani (1983), the factors triggering DOM were primarily considered to be noun class and humanness of the DO-referent, illustrated by examples that were restricted to human proper names, kinship terms and strong personal pronouns (mostly). From the 1980s, the incompatibility of the DOM-marker *à* and the definite article came under scrutiny (cf. Damiani 1983 and Marcellesi 1983), followed by a first detailed study by Marcellesi (1986).

The description of Corsican DOM in the literature is based on limited data. Many statements are backed up by isolated sentences of unknown origin. Nevertheless, authors mention animacy, noun class(es), and the incompatibility of the definite article and the *a*-marker as the main characteristics of Corsican DOM.

Humanness is one driving force for the application of DOM in Corsican. But an explicit and systematic linkage between humanness and single noun classes is still missing. Therefore, it is necessary to elicit data for missing or underrepresented noun classes, to allow for more general statements. Sentences with DO$_{[human]}$ representing *all* relevant noun classes (as mentioned in Chapter 2) must be tested. Among these DOs it is possible to also distinguish items that are lexically specified for humanness, e.g. certain indefinite proforms such as *ognunu,a* 'everybody', exclusively used with a human referents, versus underspecified items such as *tutti* 'all'. Is there any observable systematic difference ?

The second and most persuasive factor is the presence of a definite article functioning as a 'preventer' of *a*-marking. Therefore, DOs of different noun classes that allow for definite articles must be systematically checked with appropriate examples. To explore if the definite article generally blocks the *a*-marker and which option(s) are preferred, examples with a definite article in combination with the *a*-marker or even without any of these two elements must also be tested. Beyond the definite article, other prenominal elements, such as demonstratives, quantifiers, or indefinites, need to be included in the investigation, to explore the extent to which they block (or not) the appearance of the *a*-marker. The data are too scarce to support all the statements regarding DOM and the definite article.

This dissertation presents new data regarding the specific cases of indefinite pronouns, as well as the aforementioned distinction between those items lexically specified for

humanness and those not specified (i.e. *ognunu,a*[human] 'everybody' versus *tutti*[±human] 'all'), and their functional distinction between proform-only and proform-and-determiner function (e.g. ningun/o,a[±proform] and *nimu*[proform]).

Three research questions can be derived:

1. How is humanness related to noun classes and syntactic structure, considering the *a*-marking of many DOs[human], but not all?

2. Why do the definite article and the *a*-marker not co-appear? (Why) does the definite article prevent or block *a*-marking? How can is this distribution be explained in semantic–interpretative and syntactic–distributional terms?

3. And finally, the DOM-patterns of certain lexical items such as indefinite and negative quantifiers fall outside the rules developed for other lexical patterns, both with regard to semantic and/or syntactic argumentation. By what means may they be accounted for?

The investigation focusses on the semantic and syntactic structure of the Corsican DP. Other semantic and syntactic factors known to influence the application of DOM in other (Romance) language varieties will be considered, but only marginally, such as verb semantics or mode, raising effects of the DO nominal, or more generally word order divergent from the basic order.

To answer these research questions, the issue of quality and quantity of data must be adressed first. Then, the descriptions of Corsican DOM as published in the literature may be tested against the newly gathered data. These goals may be summerised thus:

1. *Quantity and quality of sentences*:
    a) to increase the number of sample sentences of the kind already presented in the literature, particularly in cases where descriptions are based on isolated examples
    b) to include new sample sentences containing semantic features and syntactic structures that are currently under- or not represented in the literature
2. *Corsican DOM in the literature*:
    - to use the new data to test existing descriptions of the phenomenon of DOM in Corsican

Previous studies on Corsican (DOM), as presented here in Chapter 4, and the methods applied therein, do not provide a large and diverse enough database to answer these questions. My earlier work on Corsican DOM (cf. Neuburger 2008), based on a corpus of written data, shows that DOM in Corsican occurs at a very low frequency. Gathering enough examples by collecting written material and/or naturally spoken language is not possible in a limited timeframe, partly due to specific circumstances in Corsica; for details see section 5.4).

For this thesis, I conducted a controlled study and collected data with a written questionnaire. I presented Corsican sentences which varied with regard to the presence of a definite article, *a*-marking, other prenominal items, noun classes and animacy values. I asked informants to choose the option they perceived as 'good' or 'best' and to evaluate the other options, i.e. I applied a combination of forced choice and evaluation (for the method in detail, see Chapter 5). The questionnaire allowed for direct comparison with the examples in the literature, and the enlargement of the database for those noun classes that were underrepresented or not considered at all.

5 Methods and fieldwork

For my fieldwork on site I worked with methodologies known as 'forced choice' and 'judgement' or 'evaluation', based on a written questionnaire. With regard to Corsican DOM, my primary objective is to understand the structure of noun phrases which function as DOs and the implications given by these structures. Beyond these the specific semantic issues, structural characteristics, i.e. syntax, are also of interest to me.

There is no existing annotated digital database of naturally-produced data for Corsican which would provide a sufficiently large and varied sample of DOM-data for the present project. In addition, the phenomenon of Corsican DOM occurs at a very low frequency in naturally-produced language, as encountered in Neuburger (2008) on the basis of a corpus with written texts. The forced choice paradigm offers the opportunity to develop a robust database, for observation, testing and analysis of DOM patterns regarding the DO-structure.

The following chapter gives a brief introduction to the research on judgement data (section 5.1), and the method chosen to collect empirical data (section 5.2). In section 5.3 I present my questionnaire and its handling. I then fo on to talk about general issues regarding fieldwork in Corsica and challenges faced by previous linguists working on-site (section 5.4.1), before presenting my own fieldwork in section 5.4.2. The chapter concludes with a summary in section 5.4.3.

5.1 Grammaticality and acceptability judgements

Many syntacticians (in Generative Grammar) use judgement data by L1-speakers as a reliable source and basis for their theoretical statements (cf. Featherston 2007: 271). A detailed historical discussion on this topic is beyond the scope of this section. Instead, the following paragraphs discuss the criteria that must be applied, in the collection of judgement data for analytical purposes, to ensure that the resulting database is reliable. These criteria, which also apply to my own data collection process, are discussed in the following paragraphs.

To gather judgement data speakers are asked to evaluate sentences in terms of their 'grammaticality' and 'acceptability'. In the 1990s, a lively discussion on the methodology and the psychological aspects of grammaticality judgements was reopened by two monographs, published by Carson T. Schütze (1996) and Wayne Cowart (1997).

Schütze (1996) provides a detailed insight to the then-current state of the art about grammaticality judgements in linguistics. He emphasises the shortcomings of the method, in

particular the absence of a control of random factors that influence the linguistic judgement process. The basis for Schütze's (1996) theoretical approach is Chomsky's Competence and Performance model, which assumes that grammaticality is a competence notion, while acceptability is a performance notion. Following this line, the naïve non-linguist speaker is not able to judge grammaticality, as grammaticality is not accessible to her/him; any judgement conducted by such an informant will inevitably be an acceptability judgement (cf. Keller 1998: 2, fn.1). In addition, "[t]he goal of linguistic theory, under this view, is to describe the knowledge, independently of (and logically prior to) any attempt to describe the role that this knowledge plays in the production, understanding, or judgment of language" (Schütze 1996: 20).

Schütze (1996) lists several factors, that affect data collection and he distinguishes task-related factors from subject-related factors. The former are procedural factors, such as measurement scales (nominal vs. ordinal vs. interval scales), instructions given to the informant beforehand, mode and order of presentation and stimulus factors, as well as content, meaning or lexeme-frequency. Subject-related factors include field (in)dependency[63], linguistic training, and handedness. Taking into consideration the psychological impacts that result from these factors, Schütze (1996) develops a psychological model of the judgement process (cf. Schütze 1996: 171–183) from which he derives his methodological proposals regarding materials and procedure (cf. Schütze 1996: 183–200) (see Table 12):

---

[63] Field (in)dependency: "A field dependent (FD) person fuses aspects of the world and experiences it globally, whereas a field independent (FI) person is analytical, differentiating information and experiences into components" (Schütze 1996: 177).

Table 12: Proposed **methodology for judgments**, by Schütze (1996: 183–200)

*Regarding **materials**:*	*Regarding **procedure**:*
1. **Order of presentation**: randomising the order of the material 2. **(Un)Grammaticality**: an even balance between grammatical and ungrammatical sentences 3. **Lexicalisation**: different lexemes for each sentence type 4. **Context**: contextualising the sentence under investigation 5. **Processing**: excluding sentences that cause processing problems 6. **Results**: minimal pairs to gather fine-grained results	1. **Kind of informants**: this is selected according to the aims of the investigation: i.e. if the competence of a "normal native speaker" is under investigation, the informant should *be* a "normal native speaker" (Schütze 1996:186); linguists should be excluded to avoid theoretical bias 2. **Number of informants**: large enough to carry out statistical tests; if a specific issue such as dialectal variation is under investigation, informants should represent speakers of various dialects 3. **Individual differences**: record on a questionnaire 4. **Instructions**: specific instructions on how to accomplish the survey 5. **Rating scale**: type depends on the issue under investigation 6. **Warm-up trials**: to ensure understanding of the procedure by the informants 7. **Filler sentences**: sufficient in number, of any type

Even if all details are taken into account, variance occurs. Cowart (1997), who recommends a strategy similar to the methodology of Schütze (1996), compares this to the variance seen in the field of behavioural data, i.e. random variance around a stable mean, which is observed in nearly all scientific measurement.

While Schütze (1996) provides a more theoretical approach to the topic, Cowart (1997) focuses on the practical implementation of grammaticality judgements in linguistic fieldwork; the preliminary theoretical considerations and practical instructions, and the data collection and the evaluation of the result, as exemplified in various surveys.

Thus, Cowart (1997) amplifies the relationship between objective methods and their application to subjective phenomena/evaluation, i.e. speakers' judgements on language data. To close the gap between the objective and subjective value system, he follows the insights of philosophers of science: "so-called objective observations were themselves subjective experiences, not fundamentally different than any other sort of subjective experience" (Cowart 1997: ix).[64] Hence, the difference between *objective* and *subjective* does not depend on the object of observation nor the individual that observes. What classifies an observation as *objective* is the inter-subjective agreement we can achieve on it, i.e. "the observations where different observers can reliably agree […]" (Cowart 1997: ix). To justify sentence judgements

---

[64] Cowart (1997) refers here to Popper (1959).

as an appropriate method to gather *objective* results on language data, Cowart presumes that:

> each speaker's I-language can be seen as an estimate of the grammatical norms of
> the community (or communities) within which the individual acquired the language
> and currently uses it, although this estimate is obviously subject to distortion by a
> variety of factors operating on the acquisition process and the generally slower and
> perhaps less orderly processes of grammatical changes in adults. (Cowart 1997: x)

Every speaker learns and knows her/his mother tongue, even if variation caused by internal or external factors might be observable when in use (i.e. "any one person's response to any one sentence is usually massively confounded" Cowart 1997: 47). Nevertheless, this starting point allows us to use speakers' judgements as an inter-subjective method. Thus, it is crucial to obtain:

> […] the standard or average pattern of judgments across sets of related sentences
> within a coherent linguistic community. By asking for judgments from many
> individuals, we can uncover typical patterns of judgments within communities;
> these patterns should bear on theories of the grammars that shape those judgments.
> Variants of these methods may also be useful in providing reliable descriptions of
> the patterns of response typical of individual speakers […]. (Cowart 1997: x)

The sought-after stable mean, then, is the average pattern of judgements which evolves from the combination of a *set of related sentences* and *judgements of many individuals.* This means it is necessary, "to test multiple exemplars of each factor combination, assuming that they vary randomly within each such cell of the design, to see if the effects of the theoretically interesting factors stand out over this random noise" (Myers 2009: 427). Or as Featherston notes, "it follows that the only reliable source of perceived well-formed evidence is the judgments of a group, in which the slips and noise cancel each other out" (Featherston 2007: 289).

When relating the average pattern of a set of related sentences to grammar, it is important to mention the position taken by Cowart (1997). He "assume[s] that theories of grammar are *partial* models of the human cognitive system that should, among other things, help to explain why judgements pattern as they do" (Cowart 1997: 7; emphasis: KAN), and that "[…] not *all* of the variation in judgments is controlled by the syntactic properties of the target sentences but that *some* is" (Cowart 1997: 10, 11; emphasis: Cowart). He is aware of the fact that grammar only partially explains the patterns of judgement data. Two questions thus remain

open. First, "is […] there any aspect of grammatical theory that can or should constrain the procedures by which judgment data are gathered or analysed" (Cowart 1997: 7)? And second (not mentioned by Cowart in this instance), other facts which explain how judgements pattern – or not pattern –, i.e. extra-linguistic factors such as missing language competence due to a linguistic community's specific circumstances, or the effects of language planning or language contact phenomena.

To summarise: if we work with acceptability judgements we have to be aware that we are asking speakers about 'their opinion' on language, and we will not gather direct information about language competence. Furthermore, judgements will reveal how a specific form is perceived and evaluated, but will not show how often, and under which conditions, it is used (cf. Boberg 2013: 134; following Labov 1972). Boberg (2013) estimates the benefits of judgements in the context of morphology and syntax as rather scant, if gathered using on the basis of a written questionnaire, since too many factors influence the speaker's choice. For example, if speakers are instructed to choose only one solution among different options, intra-speaker variation might be covered. If sentences are presented in a list, informants may correlate answers to previously given examples. These confounding factors can be counteracted to a certain degree by the methods proposed by Schütze (1996) (see Table 12, above).

What makes surveys based on a written questionnaire quite useful, nevertheless, is the relative usability of the approach, the ability to prepare the questionnaire in detail in advance, the uniformity of survey copies, and the ability to gather a large quantity of data (cf. Boberg 2013:140). While the aim of the questionnaire is to obtain the required quantity of data, the feasibility of implementing the survey is also an important consideration.

5.2 Forced choice and evaluation

I opted to utilise a modified 'forced choice evaluation' paradigm, with a focus on forced choice: primarily, informants were presented with several options and asked to choose which option was the best; secondarily, if possible, they were also asked to evaluate the other options. This approach offers the advantage of comparing two or more conditions: it "directly answer[s] the qualitative question *is there a difference between these conditions?*" (Schütze & Sprouse 2014: 33). Among its limitations, Schütze and Sprouse note first that the answers do not "provide information about the *size of the difference* between conditions, in the form of the proportion

of responses" (Schütze & Sprouse 2014: 34); second, as informants only *choose* between given sentences but does not evaluate who well the single sentences are perceived, it is not clear – assuming acceptability to be a matter of degree – "where a given sentence stands on the overall scale of acceptability" (Schütze & Sprouse 2014: 34).

In the first round of the survey, the forced choice task allowed for the testing of examples from the literature, and from Neuburger (2008), including previously underrepresented noun classes, to learn out more about the semantic and syntactic conditions for Corsican DOM. For the second round, informants were asked which option(s) they would exclude as 'bad', and on which sentences they were 'undecided/uncertain' (represented as '?'). Informants could also skip sentences where they did not know how/what to answer (represented as 'ø' in the tables in the appendix, Chapter 9). The detailed distribution of answers given by informants is shown in the tables in the appendix (Chapter 9), sorted by the nominal classes considered in the survey.

Unlike electronic or online questionnaires, where a forced answer can be implemented electronically (e.g. 'no answer' inhibits continuation), paper-and-pencil questionnaires allow for incompleteness (e.g. it is possible to 'forget' answers, if the inquiry is not supervised). Some of the questionnaires for my survey were completed in my absence, for example in class at university, or by family members of informants not present at the moment of investigation. In these instances, I had no influence on how the questionnaires were completed. This resulted in diverging numbers of answers for individual sentences; for example I may have received 48 answers for one sentence, but only 46 answers for the next sentence.

5.3 Questionnaire

As noted at the end of section 4.8, this investigation focused on the semantic and syntactic structure of the Corsican NP in DO-function. To this end, I looked at proforms, full lexical nouns, and proper nouns, as potential triggers for DOM. I distinguished between different subclasses of proforms, such as personal pronouns, interrogatives, quantifiers, negative and indefinite pronouns, demonstratives and relative pronouns. I also considered the different animacy values, [human+] (for religious figures), [human], [animate] and [inanimate], and whether the respective proforms are lexically specified for animacy, (e.g. *nimu* 'nobody'), or not (e.g. *tutti* 'all'). Furthermore, I considered the ways in which different lexical items are used – whether as proforms only (e.g. *qualchissia* 'someone') or taking a prenominal determiner function (e.g. *qualchunu,a* 'some').

Table 13 gives an overview of the above-mentioned triggers and their respective lexical

items.

Table 13: Overview of the DOM-triggers and lexical items in DO-function (as used in the QIab survey)

noun class	animacy value	lexical items	comments
proforms			
personal pronoun	lex. specified human	*ellu*	
interrogative	lex. specified human	*quale*	
interrogative	inanimate	*chì*	
universal quantifier	underspecified for animacy	*tutti*	proform and determiner function
universal quantifier	lex. specified human	*ognunu,a*	proform and determiner function
negative pronoun	lex. specified human	*nimu*	proform only
negative pronoun	lex. specified human	*nisunu,a*	proform and determiner function
indefinite pronoun	lex. specified human	*qualchissia, qualchidunu*	proform only
indefinite pronoun	lex. specified human	*qualcunu,a*	proform and determiner function
indefinite pronoun	lex. specified human and animate	*unu*	
demonstrative	underspecified for animacy	*quistu, quillu*	proform and determiner function
relative	underspecified for animacy	*chi, chì*	
proper nouns			
kinship term	human	*fratellu, babbu, moglie mammata, babbitu*	
proper noun	human+	*A Madonna, Ghjesu Cristu, Diu*	
proper noun	human	*Battistu*	
proper noun	human	*Goethe*	metonymically used
proper noun	animate	*Lione, U Sciroccu, Bella*	with/without definite article
proper noun (toponym)	inanimate	*Berlinu, Corti, A Marana, Portivechju, Bastia*	with/without definite article
proper noun (car brand name)	inanimate	*Fiat, Peugot*	
proper noun (brand name detergent)	inanimate	*Persil*	
proper noun (year, month, weekday)	inanimate	*2011, settembre, luni*	
common nouns			

common noun	lex. specified human	*omu, spiculatori, parente*	
common noun (abstract)	inanimate	*mafia*	

The choice of which sentences to include and which triggers to check was predominantly influenced by results of previous studies. Although the literature on Corsican DOM is vast, the databases underpinning its constituent studies are vague and inconsistent; statements are based on single examples of unknown origin, and potential triggers, such as humanness, are only included indirectly, by the choice of lexical items, and not explicitly referred to. Thus, while the majority of sentences chosen for inclusion in the questionnaire where taken from the literature, further examples were added, to supplement the range of different noun phrases or different combinations of feature values of the respective noun phrase in DO-function. Known DOM-triggers (i.e. the semantic and syntactic structure or the noun phrase in DO function, the animacy value of the DO-referent, the noun class(es) with which the DO is associated, and the syntactic structure of the DO-DP) were systematically tested.

Table 13 shows an overview of the lexical items (with the relevant semantic and syntactic structure) that were testes in terms of DO-function. This included the systematic consideration of the triggers, *animacy*, and *noun classes*.

With regard to animacy: I first distinguished between the four values of [human+] (for saints and God), [human], [animate], and [inanimate]; then, I tested metonymically used human names (e.g. *Goethe,* referring to the written work of the author); in addition, I distinguished between lexically specified items for humanness (e.g. *ognunu,a* 'everybody'), and items that may refer to animates and inanimates alike (e.g. *tutti* 'all').

With regard to the association of the DO with certain noun classes, I made an initial distinction between proforms, proper nouns and common nouns. Then, within the class of proforms, I further considered differences with regard to functional variance; some items may exclusively be used as proform (e.g. *nimu* 'nobody'), while others appear in determiner function as well, i.e. in prenominal position (e.g. *nisciun'omu* 'no.one_man').

With regard to the syntactic structure of the DO-DP, besides syntactic differences between proforms, proper nouns and common nouns (i.e. the internal structure [DP [NP]], see section 2.3.1 for more detail), I also tested different lexial items in prenominal position – mainly the definite article, but also demonstratives, and negative and universal quantifiers.

Among the noun classes with which the DO is associated, proper nouns play an important role

137

in Corsican DOM; their overt *a*-marking cannot be unambiguously attributed to the degree of animacy of the referent. Proper nouns that refer to inanimate objects were also included in the survey, such as metonymically used proper nouns (see example (38) below).

(38) Iab.35 *Leghju*   *à/ø*   *Goethe*

            Read.1SG   DOM/Ø   Goethe

            'I read [(a) book(s) written by] Goethe'

The same holds for brand names, such as the car brands *Fiat* and *Peugot* (see the Questionnaire QIab, examples numbers 24 and 27 (i.e. Iab.24 and Iab.27) in the appendix, Chapter 9), or *Persil*, a laundry detergent powder (see Iab.30, appendix, Chapter 9). In the context of toponyms, the influence of verb semantics in terms of affectedness was also tested (see example (39)/Iab.32 'to cover', versus example (40)/Iab.45 'to destroy'):

(39) Iab.32 *U*   *fume*   *copria*   *à/ø*   *Portivechju*

            The   smoke   covered.3SG   DOM/Ø Portivechju

            'The smoke covered Portivechju'

(40) Iab.45 *U*   *focu*   *distruhje*   *à/ø*   *Corti*

            The   fire   destroys.3SG   DOM/Ø Corti

            'The fire destroys Corti'

As well as at the semantic nature of the DO-referent, this study also looks at the syntactic structure of proper noun-DP, in particular the way in which definite articles influence the occurrence of DOM (Iab.13, Iab.14, Iab.33, and Iab.34).

      This study also focused on proforms in DO-function. The importance of humanness of the DO-element has not been explicitly investigated with regard to the implementation of DOM in the context of proforms. The role of definiteness was investigated in this study by means of including indefinite pronouns, as well as negative and universal quantifiers. Here too, the study went beyond the semantics of the DO-referent, to test these element in both pronominal and prenominal positions (where they have an adjectival function); see examples (41)/Iab.23, (42)/Iab.05), (43)/Iab.43 and (44)/Iab.44:

(41) Iab.23 *Vecu   à/ø   qualchidunu*

see.1SG   DOM/Ø   qualchidunu

'I see someone'

(42) Iab.05 *Cunnosci   à/ø   unu   chì sa…*

Know.2SG   DOM/Ø someone   who knows.3SG

'Do you know someone who knows […]'

(43) Iab.43 *Ùn   vegu   à/ø   nisciun'omu*

Not   see.1SG DOM/Ø no.one_man

'I do not see no man/I don't see anybody'

(44) Iab.44 *Cunnoscu   à/ø   iss'omu*

Know.1SG   DOM/Ø this_man.

'I know this man'

Note that all these examples were carried over from the literature without any changes to the orthography, with the exception of the accent on the DOM-marker, *à,* which was added in the questionnaire when not present in the literature. Spelling differences according to diatopical reasons were maintained.

In addition to details of the questionnaire design, some technical information regarding questionnaire nomenclature must also be mentioned. The questionnaire name, (e.g. QIab) was derived as follows: first, the questionnaire identity is denoted by the letter 'Q' for 'Questionnaire' accompanied by a roman numeral ('I', 'II', etc.); then, the small letters 'a', 'b', or 'ab', indicate whether the questionnaire was used for the first field study, survey (a), in June 2010, and my second field study, survey (b), in September 2011. The examples from the questionnaires follow the same system; for instance, the seventh example from the first questionnaire, used in both surveys, is labelled QIab.07, or simply Iab.07.

In accordance with the recommendations of the *Eurotyp*-project guidelines,[65] the following metadata were also elicited for each informant: name, gender, profession, degree of competence in the language under investigation, and languages spoken by the informant other than Corsican. I included additional questions on the personal background of the informants, requesting details about the language behaviour of their parents and about residences (place of birth and current residence), as well as lasting absences (duration, location) from Corsica.

Each informant was informed in writing that all personal and linguistic data would be anonymised and that the data would be exclusively used for scientific purposes, including the publication of this thesis. The informants declared their agreement by signature. Every informant received the same questionnaire.

The questionnaire included examples that followed expected patterns based on the existing research literature, as well as examples that contradicted those patterns. Thus, *a*-marked examples were presented both where *a*-markingwould be expected to be rated as 'good', and where *a*-marking would be expected to be rated as 'bad'. And likewise, examples without marking were presented both where marking would and would not be expected (based on the information given in the literature). In addition, the survey included filler sentences.

In accordance with Schütze's (1996) recommendations regarding 'materials' (see Table 12, above), sample sentences with the same nominal categories were randomised, the number of grammatical and ungrammatical sentences was balanced (always pairing a grammatical and an ungrammatical option), and DOs from the same noun class were embedded in lexically different sentences.[66] For most sentences a context was given, and sentences were structured in a approachable manner (by using instead a simple lexis), to prevent processing problems.

5.4 Fieldwork

The unique conditions encountered in Corsica influenced the fieldwork. Also other linguists described difficulties during fieldwork. In many studies, the (social and cultural) circumstances of empirical data collection are not published. This is regrettable, because better knowledge of the site could facilitate fieldwork for future researchers. Therefore, I have included a section regarding site-specific issues and the experience of other linguists, before presenting my own

---

[65] https://www.eva.mpg.de/lingua/tools-at-lingboard/questionnaire/eurotyp-guidelines.php
(last accessed 13.08.2020)
[66] To minimise the "lexis-specific and content-specific effects" (Featherston 2007: 291), DPs of the same noun class were lexicalised differently for the sample sentences. This also has a bearing on the impacts caused by presenting continuously sentences of similar structure to the informants: to alternate the lexical content reduces the effects of mental fatigue, irritation and incomprehension (cf. Featherston 2007: 292).

fieldwork. The two fieldtrips will be considered separately, because they proceeded differently. As I used the same questionnaire twice, the collected data result in one single corpus.

5.4.1 *Field research in Corsica as described in previous studies*

The current Corsican linguistic reality appears – e.g. in the media, in scientific work, in direct contact with Corsicans and Corsophones – to be very heterogeneous. The linguistic landscape consists of many intra- and interspeaker varieties, diversified not only according to diatopic features, but also according to the degree of language contact, the educational background of the speakers (e.g. language learning through school and university and the teacher's background), and the personal circumstances of the speakers (e.g. profession, migration, and leisure activities). To accomplish a controlled experimental investigation with a small number of informants, dealing with a high degree of language variation (regarding diatopy and language competence) and a scarcity of monolingual L1 speakers, is a challenge.

The procedure and results of any given survey, or the handling of data gathered (and elaborated), are influenced not only by the informants' language competence and behaviour, but also by the investigator's (language) background, the situational circumstances, and the methodology and techniques applied.

In the following paragraphs, I briefly summarise the experiences of linguists who have worked in Corsica and collected empirical data.

When working for the NALC-project (see section 3.1.5 for details) in the late 1970s, Mathée Giacomo-Marcellesi (1978b) and her team were confronted with unexpected and inconsistent language behaviour from the informants. The interviews were conducted in everyday situations where Corsican was the predominant medium of communication. All the interviewers were L1 speakers of Corsican, as were the informants. Despite this, Giacomo-Marcellesi and her team found the speakers to be reticent:

> Dans cette situation, où nous trouvons au milieu de gens dont le corse est le moyen de communication quotidien, courant, naturel, nécessaire dans le rapports de sociabilité et d'échange, le corse nous monte normalement aux lèvres, bien que tous soient capables d'utiliser le français, et aient même tendance à le faire devant la 'frusteri' que nous sommes. (Giacomo-Marcellesi 1978b: 147)

> [In this situation we find ourselves in an environment where Corsican is the

common, current and natural medium of communication in social relations and social exchange, Corsican flows naturally from the lips, although everybody is able to speak French – and so they did in front of us strangers; KAN].

By way of example, she mentions a situation where the interviewing team was introduced by a third party who spoke French. The informant, a fluent L1-speaker of Corsican, did not speak any Corsican until the third party left. This is an extreme case, but other, similar situations arose (cf. Giacomo-Marcellesi 1978b: 147). Hence, being in a French-speaking context may inhibit even those competent in Corsican from actually speaking it.

Giacomo-Marcellesi and her team concentrated on the older generation, because the topic of the survey was the traditional knowledge about manufacturing processes and forms of (agricultural) life. The older generation's general knowledge on the topic, and their specific language in particular, were considered 'richer'. But when surveying older informants, the investigator faced the problem that these informants often were circumlocutory and their answers digressive. As the project consisted of eliciting specific terminology, this made it difficult to elicit the term or information of interest. Collaboration with the younger generation, while less detailed regarding the language itself, was more target-oriented. By way of a solution, Giacomo-Marcellesi and her team tried to conduct the survey, whenever possible, in the presence of two generations (cf. Giacomo-Marcellesi 1978b: 147).

For their project on the diatopical variation of Corsican[67] (cf. Thiers 1986; see also section 3.1.5) Ghjacumu Thiers and his team were also confronted with problems relating to the recruitement of informants suspecting a political motivation behind the survey: "Quant aux refuse de l'enquête, ils ont parfois été circonstanciés par des réponses où transparaît une hostilité plus ou moins grand à l'encontre d'une investigation parfois considérée comme un moyen de propagande de l'idée nationalitaire corse" [As for the refusals of the survey, they were sometimes detailed by responses showing a greater or lesser hostility to an investigation that was sometimes considered a means of propaganda for the Corsican nationalist idea; KAN] (Thiers 1986: 70). Furthermore, some institutions or groups as a whole declined to participate citing the topic of Corsican nationalism, with which they did not want to be associated. Any linguists conducting fieldwork on Corsica has to deal with reservations of this kind, either directly (i.e. people do not want to participate) or indirectly (i.e. people bring up the topic at

---

[67] Note that the project was based on a collection of texts of different diatopical Corsican varieties, translated from a French source text, to be presented to informants as audio-data so as to garner their opinions on diatopical variation.

least).

For a non-Corsophone linguist undertaking empirical fieldwork in Corsica without a pre-existing, reliable network of informants, gaining 'access' to Corsophones can prove difficult. To find a potential informant, i.e. a speaker of Corsican, it is essential to ask questions on language competence ('Do you speak/understand/etc. Corsican?', 'Do you know someone who knows Corsican?', etc.). But, as Blackwood (2004b; 2008) experienced during his surveys (see also section 3.1.5), self-reported and unverified answers regarding language competence have to be treated with caution (cf. Blackwood 2004b: 250, 251; 2008: 107, referring thereby to Labov 1972; Trudgill 1974; Milroy & Milroy 1985); informants may exaggerate or undervalue their own practices and competences. Moreover, the question "Do you speak/understand Corsican?" does not ask the informant to specify their degree of language competence, which may range from the ability to greet and understand some idiomatic expressions, to the ability to handle technical discussions about regional policy. To avoid the problems of over- or underestimation in self-evaluation of language competence, Blackwood conducted his second survey in 2000 without asking the informants about their language competence (cf. Blackwood 2004b: 250).

As well as difficulties relating to informants, methodological questions also arise, when conducting fieldwork, and when analysing and making the data available. Chiorboli (1987) questions the process of selection of the varieties for diatopical studies; as for the survey he was involved, certain varieties were included, while other varieties were excluded. But this selection remained without further explanations. Additionally, using translations (e.g. from a French source) introduces the risk of interference and hypercorrection (cf. Chiorboli 1987: 77).

Blackwood (2004b), when analysing the data from his surveys on language elaboration from 1999 and 2000,[68] became aware of certains disadvantages of his method; asking for isolated lexemes without a context was perceived as unnatural by informants and suggested a 'test situation', especially when they were being recorded. Blackwood deduces that this perceived 'test sitation', and the recording, could both have influenced informants' answers (although he does not further specify how). In addition, the number of lexemes – three nouns in the 1999-survey – is too small to draw any conclusions about the elaboration process. For

---

[68] Note that the 1999-survey (cf. Blackwood 2004b) was based on three lexical items, 'aeroplane', 'difference' and 'subscription', to be translated into Corsican. The goal of the survey was, to evaluate the language-elaboration process and to reveal which Corsican lexemes are given by the informants: the 'distantiated' version, following a puristic elaboration approach or the 'polysomic/galliced' version, based on language contact with French (see also sections 3.1.3 and 3.1.5).

the survey from 2000, he augmented the number of lexemes (cf. Blackwood 2004b: 238, see also section 3.1.5).

But even if linguistic data is successfully gathered, as for the NALC–BDLC project, making it available to a broader public can be difficult; the BDLC, designed as an online database, was often offline during the course of its development, sometimes for months, and if online was often very unstable. Many of the aforementioned aims of the BDLC (for details, see section 3.1.5) are still in development and are not available in the online version. In particular, systematic annotation of the data, allowing analyses from different linguistic approaches (phonetic, morphological, etymological, etc.), is only possible to a very limited extent.

To summarise, fieldwork on Corsica depends, to a high degree, not only on the informants' language competence and behaviour, but also on the linguistic researcher conducting the survey. As we have seen, Corsicans can react quite cautiously to foreigners in their 'natural surroundings', whether or not the researchers are also Corsophones (cf. Giacomo-Marcellesi 1978b). Sometimes this is due to historical–political reasons, such as the link between (right-oriented) politicking and language activism (cf. Thiers 1986). Additional problems arise if the researcher her/himself is not Corsophone (cf. Blackwood 2004b) and has to operate without a reliable network of informants. Then, the project depends on the contacts the researcher makes and the informants' self-evaluation of their linguistic competence, the latter of which also may influence the survey results. Methodological factors, such as sample size and data collection (cf. Chiorboli 1987; Blackwood 2004b) or data processing (as in the BDLC-project) also influence the data that are gathered during fieldwork and available for analysis.

5.4.2 *Fieldwork for this study*

In order to gather an appropriate data-set, to address the research questions posed in section 4.8, two fieldtrips were planned and conducted. Preliminary considerations concerned the questionnaire design (as described in section 5.3 above) and the organisation of the fieldwork with regard to *informants*, *data* and *locations*, as detailed below.

In line with Featherston's (2007) and Myers' (2009) suggestions for dealing with linguistic data I tried to make the whole process of data collection – i.e. questionnaire methodology (as described above), and survey preparation and execution(with regard to informants and the elicitation process) – as transparent as possible.

Survey (*a*): June 2010

Prior to commencing my fieldwork on Corsica, I searched for (a) Corsican speaker(s) to develop the questionnaire and to conduct a first trial of the methodology. To this end, I tried to contact Corsicans living abroad – with no success.

Only very few linguists in Germany and Switzerland have been working on Corsican during the last decades. They have been concerned with research questions based on written language documentation (e.g. Fabellini 2010) or have been working in other domains of linguistics (socio- or media-linguistic studies, i.e. Kailuweit 2015).

Before the fieldtrip, I identified institutions, organisations or private persons on Corsica with an internet presence, which seemed likely, in the broadest sense, to enable contact with Corsophones. Out of approximately seventy emails sent, approximately sixty received no reply. I was, however, able to make some appointments: in Bastia, with the public library, a cultural institution (offering theatre, dance, traditional song, and Corsican language classes), and a radio-station; in Aiacciu, with a publishing house; in Cervione (Castagniccia region), with a cultural institution; and in the Balagna region, where I met an author.

The planning for my fieldtrip (to take place in June 2010) took into consideration organisational issues mentioned above, namely informants, data, and locations.

INFORMANTS: FINDING AND INTERVIEWING 30 ADULT L1-SPEAKERS OF CORSICAN

This survey was planned to consist of 15 women and 15 men, with 5 representatives for each age group (18–36/37–66/67+ years). The division according to gender was motivated by the different linguistic behaviour of women and men, as described in section 3.1.5. The classification according to age was motivated by the Corsican historical–cultural circumstances (see section 3.2 for historical background) whereby informants aged 18–36 have grown up (in addition to language acquisition – or lack thereof – at home) with Corsican language teaching through public education, while the generation aged between 37 and 66 years is considered to be part of the so-called 'lost generation' with scarce or no language acquisition of Corsican at a young age, within the family or in school. This 'lost generation' have been involved in the political processes of language planning since the 1970s and 1980s. The older generation, aged

67+, have acquired Corsican, at least partially, within the family.[69]

With regard to profession or educational background of the informants, I did not apply any conditions in advance.

DATA: FORCED CHOICE AND EVALUATION, BASED ON WRITTEN QUESTIONNAIRE, FREE SPEECH RECORDINGS, ADDITIONAL COLLECTION OF AUDIO- AND PRINT DATA

Along with the questionnaire, recordings based on free speech conversations between interviewer and informants were planned. To enlarge the database, I aimed to collect Corsican audio- and print documents, i.e. audio-documentation of radio-programs, movies, books, newspapers, and journals, as well as documents from archives and libraries.

LOCATIONS: DIATOPICAL VARIATION?

Ideally, surveys should be conducted in different locations with representatives of the local varieties. Accordingly, my surveys were planned to take place in three main locations: the cities of Bastia, Aiacciu, and Corti. These locations were chosen with regard to diatopical differences, with Bastia in the north-east and Aiacciu in the south-west representing, respectively, the northern and southern varieties. Corti was added as it hosts the Corsican university, in which environment the availability of willing informants was expected to be high. All selected locations were urban centres, where access to informants was expected to be easier.

Fieldtrip 1 (1–30 June 2010)

The structure of this section re-iterates the three organisational parameters outlined above – i.e. informants, data, and locations. In each case, I specify which of the preliminary claims were met and, where – and why – they failed.

INFORMANTS

In general, it was a problem finding *any* speakers of Corsican. From the few contacts made in advance, even fewer resulted in an appointment. On Corsica, I approached potential speakers by speaking French. Many people I talked to denied any language competence or any personal relationship with Corsophones when directly asked. Reasons for this reaction may have ranged from them genuinely not being a competent speaker, through not being interested, to the fear of being associated with any politicking (as previously mentioned by Thiers 1986: 70, section 5.4.1). The absolute number of L1-Corsophones is estimated to be quite low and the historical–

---

[69] The classification according to age is actually based on the year of birth, to unify the data of both fieldworks with regard to age: aged 67+ = born 1944 and before; aged 37–66 = born between 1954 and 1973; aged 18–36 = born between 1975 and 1992.

cultural context was known to me, so I was not surprised by the answers and behaviour of the people I addressed. At a late stage of the first fieldwork, I realised that speaking French myself made it easier for the addressed person to deny any Corsican language competence. This approach was changed for the second fieldtrip by speaking predominantly Italian (for details, see below).

As the questionnaire could not be tested in advance of the fieldtrip by an L1-speaker of Corsican, the first two informants (excluded from the survey) served as consultants for a general review of the questionnaire with regard to lexis, morphology, syntax, and sentence content (e.g. to avoid potential processing problems due to lexical choices). They were not informed of the topic of the survey. During this review process, in addition to minor corrections regarding lexis and morphology, major discussions arose regarding the questions of orthography and diatopical variation in particular.. Examples included from the literature display a certain heterogeneity with regard to the diatopical varieties cited and, consequently, so does the spelling; Corsican spelling reflects the *polinomia* concept (see section 3.1.2, above), resulting in such variations as *quillu* (northern varieties) or *quiddu* (southern varieties) 'that[demonstrative]'. This diatopical variation was also adopted for the examples I constructed myseld. By ensuring they included items from both/all varieties in approximately equal measure. The two test-informants touched upon this heterogeneity several times, being distracted by the inclusion of different varieties in one document. Even though it was clear that this could lead to problems with future informants, the approach was maintained, as all Corsican diatopical varieties should be equally represented according to the *polinomia* concept.

My initial approach – talking to people in the streets, shops, public institutions (such as cultural centres, offices and service points) – remained without major success. I had most success making contact with 'language activists' of different kinds; many of the informants were involved in one way or another in the processes of language planning, elaboration, documentation, policy, etc. None of them could be characterised as completely unbiased speakers of Corsican. Of the Corsophones interviewed, many were somehow involved in the aforementioned language processes and had reflected on the language more than a so-called naïve non-linguist speaker. I asked them too, about making contact with other potential informants – known as the 'friends-of-a-friend' technique. All these informants I have interviewed were included.

The first survey resulted in twelve participants, as shown in Table (14).[70]

Table 14: Overview of the informants, fieldwork a, June 2010 (12 informants)

Age and gender			
	18–36 years (1 informant)	37–66years (10 informants)	67+ (1 informant)
male	1	8	1
female	-	2	-
Do you currently reside in the region where you were born?			
yes	1	6	-
no	-	4	1
How many years did you live abroad?			
0 years	1	-	-
1–3 years	-	1	-
4–10 years	-	2	-
10+ years	-	7	1
Language competence (self-declaration)			
L1 Corsican	-	-	-
L1 Corsican/French bilingual	-	6	1
L1 French/Corsican bilingual	-	2	-
L2 Corsican	1	2	-

This table shows that the sample was neither balanced for age nor for gender. The group aged between 37 and 66 was well represented. This is due to the fact that I contacted various institutions where people work, so it was easiest to meet persons within this age range. Also, because I made many contacts by the 'friends-of-a-friend' technique, informants tended to recommend people of their own age group.

With regard to language competence, no informants grew up with Corsican as her/his first and only mother tongue. The informant aged 67+ grew up bilingual Corsican/French, with Corsican being the dominant language. In the middle group, two informants declared Corsican as being their L2. A general tendency can already be observed within this small sample of informants; the language competence changes from Corsican being the dominant of two L1 languages among the older generation, to a more diversified picture for the younger Corsicans. Among informants aged from 37 to 66, six of the ten declared that they had grown up in a

---

[70] For a more detailed exposition of the informants, see the appendix, Chapter (9).

bilingual setting with Corsican being the dominant language.

I focused on the completion of the questionnaire. The collection of other language data and material became secondary. This was mainly due to the limited time-frame and the sometimes lengthy process from first contact to final interview.

After being reviewed by two Corsican-speaking consultants, the questionnaire QIab was enquired with the 12 informants detailed in Table 14. All interview sessions were conducted with just the informant and myself present. Every informant was asked if the session could be recorded, but they all refused. I did not insist on recording the sessions because setting out the pros and cons for the benefit of each individual informant, to see if they would change their mind, took too much time, and I respected their refusal. I realised that judging language data was already an unnatural situation for them; the recording would have reinforced this impression (see aforementioned difficulties and comments on recordings, making the situation more "formal", by Blackwood 2004b: 237).

Sessions mostly commenced with an informal conversation about various topics. Informants were encouraged to speak Corsican, while I spoke Italian. Depending on the topic, informants switched between Corsican (sometimes also Italian) and French. Predominantly, matters relating to policy were verbalised in French, while Corsican was used when talking about private matters, such as family.

Completing the questionnaire was also accompanied by some conversation, mostly concerning sentence content, and rarely relating to linguistic issues regarding the questionnaire, diatopical differences in lexis, or the writing of proper names. One informant realised the intent of the questionnaire ("You want to know when to put the *à*, right?", m52cf.IaIb).

LOCATIONS

Only informants living in the two cities of Aiacciu and Bastia were interviewed, with the exception of one informant from a small village in the area of L'Île-Rousse (Balagna region, north-west coast). Of the 12 informants, 6 were born in or close to the city they lived in at the time of interview. The other 6 either moved across the island (2 informants) or were born abroad to parents of which at least one was Corsican.

All except one had spent some time living abroad, mostly for long periods (10–25 years), due to military service, university, and/or work.

No informant had remained settled without interruption in the location where s/he was

born. The aim to investigate diatopical variation of DOM became secondary. The approach of meeting informants in different parts of the island was still maintained, but this was more for practical reasons, as it was difficult to find any competent and willing participants anywhere on the island.

Fieldwork *b* (September 2011)

PREPARATION

INFORMANTS

The main aim of the second trip was to enlarge the group of informants, keeping in mind the intended distribution across age and gender. In particular, the aim was to increase the number of informants from the youngest and oldest generations, which had been underrepresented in the first survey, by means of contacts made at the university of Corti, and by asking pre-existing contacts for introductions to possible informants aged 66+.

DATA

The twelve informants from the first fieldtrip did not yield sufficient data. To ensure compatibility, the same questionnaire, QIab, was used again. The first survey using this questionnaire did not cause any major problems, so its original structure was retained. The goal was to get as many questionnaires filled in as possible. The search for written documentation and other language material, such as audio- or video-files, was secondary.

LOCATIONS

The plan to differentiate the data according to diatopical variation was given up because the personal histories of the informants did not allow for such an approach. All informants were included, regardless of their place of origin on the island and not only from the preselected cities Bastia, Corti, and Aiacciu.

Fieldtrip 2 (1–28 September 2011)

INFORMANTS

A total of 37 informants were surveyed, out of these 3 persons, already interviewed during the first fieldtrip in June 2010, where surveyed a second time.

With regard to initial contact, I changed my own communication strategy. Instead of presenting myself and the project in French, I started the questions and conversation always in

Italian, using the excuse that my French was not 'presentable'. This strategy turned out to be quite advantageous as I could immediately tell if my dialogue partner was able to understand/speak Corsican by her/his reaction and answer. Admittedly, this approach ran the risk that the informant might reply in Italian and not speak Corsican. But in fact it emerged as a quite reliable method of finding Corsophones, since I did not give them the option to 'hide' their language knowledge. At the same time, it enabled me to avoid the question *Do you speak/know/understand Corsican?* which self-evaluation had previously led to rather unreliable results; the actual degree of language competence remained unclear even if the speakers had to state in the questionnaire whether they have acquired Corsican as L1, in a mono-/bilingual context, as L2, etc.

Again, most contacts were made by the *friends-of-a-friend* technique; owing to the small number of participants enlisted, it was difficult to select informants according to place of origin, age, gender, and language competence.

The visit to the university was very beneficial to the task in general. Thanks to two professors, the questionnaire was completed by students of their classes, and representation of the youngest generation in the sample increased considerably.

For an overview of the informants of the second fieldtrip in September 2011; see Table 15.

Table 15: Overview of the informants, fieldwork *b*, September 2011 (37 informants)

Age and gender			
	18–36 years (19 informants)	37–66 years (13 informants)	67+ (5 informants)
male	10	8	4
female	9	5	1
Do you currently reside in the region where you have been born?			
yes	3	6	3
no	11	6	2
?	5	1	-
How many years did you live abroad?			
0 years	12	5	-
1–3 years	2	2	-
4–10 years	1	2	-
10+ years	2	4	-
?	2	-	5
Language competence (self-declaration)			
L1 Corsican	3	4	1
L1 Corsican/French bilingual	6	3	3
L1 French/Corsican bilingual	2	2	-
L2 Corsican	8	4	1

With regard to age and gender, the second survey was somewhat better balanced. It resulted in 19 informants aged between 17 (one person, included in the age-group 18-36) and 36, of whom 13 were students of the two aforementioned professors; 10 of these youngest informants were male, 9 female. This was the only age group balanced according to gender. This might be due to the high rate of female students taking the study course *Corsican language and culture*. Of the 13 informants interviewed who were aged between 37 and 66 years old, 8 were men and 5 women, while the group representing the older generation – those aged 67+ – included 1 female and 4 males. Of a total of 37 informants, 15 were women, and 22 men. The number of female informants is markedly lower than of male informants, because there were fewer women than men among 'friends-of-the-friend', and not all of these were willing to participate. I experienced more refusals from women than from men. Asked about why this might be, one female informant mentioned reticence, on the part of many women, to speak in or about Corsican to foreigners, as speaking Corsican is considered something of a private matter.

Additionally, historical–cultural circumstances (language politicking in the 1970s, 1980s) may explain the behaviour of the older women. This reasoning also explains the sometimes rather difficult process of making contact, which took many intermediate steps prior to the actual interview.

Twelve of the 37 still lived – or, better, were living once more – in the region where they were born, only 3 of the 12 having never lived abroad. Of the youngest generation, only 5 of the 19 informants had lived abroad, and only 2 of those for more than ten years. This is partly a result of changes in education. The Corsican university reopened in 1981 and is constantly enlarging the range of study courses offered. Consequently, there is no need to move to the French mainland for higher education anymore. Nevertheless, of those people aged between 18 and 36 who never left the island to live abroad, only 3 still resided in the same region where they were born. As the university is in the centre of the island and the road network does not support commuting on a daily basis, most students move to Corti or the nearby villages.

Only 6 of the 13 informants aged between 37 and 66 years oldstill lived in the region of their origin; 2 of them had never been away, 1 had only been away for two years (due to military service), while the other 3 had been away for between seven and ten years. Of the eldest informants (67+), all 5 had stayed away for comparatively long periods, of between 11 and 40 years. Consequently, the 3 who were still living in the region where they were born cannot be regarded as having had 'continuous' residency in one place.

Out of all 37 informants, four stated neither their current place of residence, nor their place of birth; a further three did not specify their place of origin. With regard to periods spent abroad, two informants did not provide any information.

Finally, with regard to language competence, 8 of the 37 informants considered themselves to have grown up monolingual Corsican until their entry into school. Surprisingly, 3 of these 8 were students, i.e. representatives of the age group that tend to be associated with monolingual French language acquisition – according to self-assessments confirmed by their professors. There were another four self-declared monolingual L1-speakers among the informants aged between 37 and 66. Among the informants aged 67+, only one was a monolingual L1-speaker of Corsican. Of the 12 bilingual speakers, with Corsican being the dominant language while growing up, 6 were from the youngest age group, 3 from the middle age group, and another 3 from the 67+ age group. Four informants were bilingual but with French having been the dominant language during infancy; 2 of these were aged between 18 and 36, and 2 between 37 and 66.

A total of 13 informants considered their Corsican to be L2; the majority of whom (8 informants) were between 18 and 36. Of the 37-to-66 year-olds, 4 had acquired Corsican as an L2, while the same was only true of 1 informant from the oldest group.

DATA

As opposed to the first fieldwork, when all informants completed the questionnaire in my presence, the dataset from the second fieldwork also included questionnaires answered by friends or family of my informants. One class of students compiled the questionnaire in a lesson after my departure. With regard to the data from the student group, it must be noted that (probably due to time limitations during class) not all students finished the questionnaire. Consequently, the completeness of answers from these informants diminishes towards the end of the questionnaire. As these were sent to me by post I only saw the results once back in Germany and could not intervene or ask the informants to complete the missing sentences.

For the second survey, three informants from the first fieldwork were interviewed for a second time while 34 informants were participating for the first time.

Regarding the completion of the questionnaire, no major difficulties arose, which was comparable to the survey conducted during the first fieldwork in June 2010.

LOCATIONS

Informants from anywhere in Corsica were accepted for the survey. As before, I worked with informants in Bastia and Aiacciu, but also interviewed university staff, students and teachers in Corti, and visited the Balagna region to meet two informants. Also, three informants in Cervione participated in the survey.

5.4.3 *Summary data and informants*

To sum up the results from both fieldtrips, a total of 46 informants participated in the surveys. Of these, three informants participated twice (i.e. once in each survey). 29 of the informants were male, 17 female. With regard to age distribution, 20 informants were between 18 and 36, representing the generation that has grown up with Corsican language classes at school. Most (13) of these had never lived abroad, which is not surprising as they were young in age (13 of them were first-year students), and had access to a broad range of higher education on the island through the university in Corti. Eleven of them no longer resided in the region where they were born. The highest number of Corsican L2 speakers (nine informants) occurred among the youngest informants, along with nine others that had grown up mono- or bilingual Corsican/French, with Corsican being the dominant L1.

154

The number of informants aged between 37 and 66 was more balanced than other age groups with regard to current residency and place of origin; 10 of the total of 21 still lived in the same region where they were born. At the same time, only 5 had never lived abroad, while 13 informants were abroad for four years and more. Reasons for living abroad were, in most cases, military service, higher education, and/or work on the French mainland. The older the informants, the longer they had tended to have lived abroad (this tendency was also seen among informants aged 67+).

The oldest generation was represented by 5 informants, 4 of which were male. Three of these lived (at the time of interview) in the region from which they originated, but all of them had been abroad for more than ten years, in 3 cases this being more than 20 years. Despite being away for such a long time, they had spoken Corsican continuously throughout their whole lives, in particular with their families. Two of them had grown up in Corsican communities of military personnel living abroad; hence, for them Corsican was not only a medium of conversation among family members, but also among neighbours and friends.

If we disregard the informants' ages, the results show that the majority of informants (23 out of 46; six did not answer the question) did not live in the regions in which they were born.

With regard to language competence, more people of the older age groups had grown up with Corsican as L1 or in a monolingual or in a bilingual context (Corsican being the dominant language for 12 informants) with respect to the younger generation, but here too, there were 6 informants who declare Corsican as their L2. Only 8 informants considered themselves monolingual L1 speakers of Corsican, another 17 having grown up bilingual with Corsican being the dominant language in childhood before entering school. Five more informants grew up bilingually with French being the dominant language.

What was common to all these informants, setting aside their self-declarations on competence, was that all informants used Corsican spontaneously and fairly frequently in different language situations. For example, Corsican was not only applied in class at university, but also at home or with friends or during certain leisure activities; if they used Corsican as a medium of conversation in a private context, I wanted to know if it went beyond idiomatic expressions, salutation or the use of borrowings.

I adapted Schütze's (1996) and Cowart's (1997) recommendations to recruit L1 speakers, for my approach. I included any candidate who spoke Corsican in a spontaneous manner, beyond the level of idiomatic expressions and the like, in at least two different contexts in their daily lives. This approach is justified in view of the *polinomia* concept (including not only

different diatopical varieties, but also varieties noticeably marked by language contact with French) and in the absence of a point of reference in terms of an official standard.

For this thesis, *current Corsican* has been under investigation. Probably not in line with traditional approaches on empirical linguistics, the informants included in the sample *do* represent the current use of Corsican in Corsica and the Corsican speaker of today, taking account of the socio-historical circumstances on the island and of the islanders. For example, if the generation born before 1960 wanted to study, they had to leave the island for at least their five years at university, Also, if male, they had to leave the island for another two years of military service. By contrast, the younger generation may remain on the island to visit university, although most of them have to move to Corti and its surrounding area. As elsewhere in the world, on Corsica too, people move into urban centres or smaller towns for economic reasons. Hence, compared to the number of those who have moved, only a vanishingly small number of Corsicans have never changed their residency and continue living in the linguistic environment in which they were born and grew up. Also, the role of language education can not be ignored; language use changes under educational approaches, even if it follows polinomic lines, as in Corsica. Also the use of language in written form changes traditionally oral languages. These are only some of the many factors that have influenced and will continue to influence the Corsican language.

The purpose of this dissertation, the morphosyntactic analysis of Corsican DOM, is based on current Corsican resulting from the linguistic reality of today, and therefore includes all data from the informants according to the criteria mentioned above.

Table 16 shows a summary of the information about all of the informants.

Table 16: Overview of fieldwork *a* and *b* with regard to all 46 informants

Age and gender			
	18–36 years (20 informants)	37–66 years (21 informants)	67+ (5 informants)
male	11	14	4
female	9	7	1
Do youcurrently reside in the region where you were born?			
yes	4	10	3
no	11	10	2
?	5	1	-
How many years did you live abroad?			
0 years	13	5	-
1–3 years	2	3	-
4–10 years	1	4	-
10+ years	2	9	-
?	2	-	5
Language competence (self-declaration)			
L1 Corsican	3	4	1
L1 Corsican/French bilingual	6	8	3
L1 French/Corsican bilingual	2	3	-
L2 Corsican	9	6	1

6 Survey Result

In the following chapter I present the results of my study, based on the questionnaire QIab and gathered during my fieldtrips in June 2010 and September 2011. For both fieldtrips, I used the same questionnaire. The results are combined to form one single corpus of QIab data, and compared to claims made about Corsican DOM in the existing literature, and to the results of Neuburger (2008) and Neuburger and Stark (2014) (see Chapter 4).

I display the data in tables, as follows: word class and (non)marking pattern are displayed in the first column; the total number of tokens recorded in the database are displayed in the in the second column; the results of the forced choice task, wherein informants had to select the 'best' option, are displayed in the third column (labelled) 'good'; the results from the evaluation task, wherein informants could indicate those options they considered to be 'bad', and those about which they were uncertain or undecided, are displayed in the fourth column (labelled 'bad') and the fifth column (labelled '?'), respectively. The evaluation task was optional. For an example, see the model table below (Table 17).

Table 17: Model table: distribution of forced choice and evaluation answers

	total	***good***	*bad*	?
à + toponym	272	**149**	27	9
ø + toponym	271	**136**	24	15

The 'total' (column two) and the sum (of columns three, four and five: 'good' + 'bad' + '?') diverge, because, while informants were required to select their preferred option (i.e. that which they considered 'good') in all cases, evaluation of the remaining alternatives was optional.

I divide the data according to the noun classes to which the respective nouns in DO-function are traditionally ascribed: first, kinship terms (section 6.1) and proper names (section 6.2), followed by common nouns (section 6.3) and proforms (section 6.4). In the subsections, where relevant, I proceed according to the order given by the hierarchies presented in section 2.1: starting from the 'left end' of the hierarchies, where definite, specific and human items are categorised; proceeding to the 'right end', where we find indefinites, non-specifics and inanimates. I also point out if there are specific lexical items for a certain animacy value (e.g.

negative quantifier *nimu* 'nobody'), or if the items are underspecified for animacy (e.g. demonstratives, *quillu, quiddu* 'this').

I introduce each noun class with what has been said about it in die literature, followed by results from Neuburger (2008). Against this background I then describe the patterns that emerge from my survey results, in terms of the distribution of responses between the available options ('good', 'bad' and '?') as displayed in the tables.

Because variances can be obscured by viewing the data in bulk, examples that diverge noticeably from the mean are singled out and discussed separately.

6.1 Kinship terms

In the literature, kinship terms are mentioned among the nouns marked by *a*. A kinship term refers to family relationsships of first and second degree, i.e. parents, children, grandchildren and grandparents, sisters and brothers, which are usually considered to bear *a*-marking. These kinship terms appear without the definite article if the persons they refer to are related to the speech participants. By implication, that means that if they are related to neither the 1st nor 2nd person, they bear the definite article (if contextually required) and *a*-marking is excluded. This is confirmed in Neuburger (2008): all 1st/2nd-person related kinship terms are *a*-marked and bear no definite article. One kinship term, *u babbu di a negritudine* 'the father of the *Négritude*', appears with a definite article and no *a*-marker.

In the literature review and the survey, I considered both the semantic relation between the referent of the kinship term and the participants, and the syntactic structure of the kinship term; this involves the application of a definite article and/or other prenominal items. In the QIab corpus, the evaluation of *a*-marked and unmarked kinship terms resulted in a less clear-cut distribution than the literature and Neuburger (2008) suggested; as shown in Table 18:

Table 18: Kinship terms

	total	***good***	*bad*	?
à + kinship term	84	**68**	2	0
ø + kinship term	84	**16**	18	1
à + possessive + kinship term	179	**84**	21	3
ø + possessive + kinship term	133	**21**	30	1
ø + def.art. + poss. + kinship term	93	**78**	6	2
Total		**267**		

The caterogy, 'Kinship terms', test sentence numbers Iab.1a–c, Iab.26a–d, Iab.29a, b, d, e, and Iab.46a and b (from the questionnaire in Chapter 9) were considered.

For each combination shown in Table 18, one example is given in Tables 19 and 20:

Table 19: Kinship term *babbu* 'dad' (QIab test sentences Iab.46a and b)

(à/ø) + kinship term *babbu*		total	***good***	*bad*	?
Iab.46a	*Ùn ai micca vistu à babbu?*	41	**34**	0	0
Iab.46b	*Ùn ai micca vistu babbu?*	41	**9**	9	0
Total			**43**		

Table 20: Kinship term *fratellu* 'brother' (QIab test sentences Iab.01a and b)

(à/ø/def.art.) + poss. + kinship term *fratellu*		total	***good***	*bad*	?
Iab.01a	*U sgiò Marcelli chjama à so fratellu*	46	**4**	10	1
Iab.01b	*U sgiò Marcelli chjama so fratellu*	46	**1**	12	1
Iab.01c	*U sgiò Marcelli chjama u so fratellu*	47	**40**	3	1
Total			**45**		

The results in Table 18 show that, when evaluating kinship terms, the majority of informants rated *a*-marking as 'good', as opposed to the numbers of 'bad' or '?' evaluated sentences. Both the '*a*-marking + kinship term' option, and the '*a*-marking + possessive + kinship term' option, were preferred over the ø-marked options, although this was less clear-cut with possessives than without (i.e. when possessives were included, informants evaluated sentences as 'bad' or '?' more often than when possessives were not included). Where informants had the option to choose 'possessive + kinship term', they favoured the option with the definite article over the *a*-marked option (see example no. Iab.01c, in Table 20).

6.2 **Proper nouns**

6.2.1 *Proper nouns [human+]: religious figures*

In the literature, only a few authors mention proper nouns of religious figures. Bottiglioni (1933–1942) includes the names of two religious figures – Jesus Christ, and the Madonna – as an extra subclass. According to him, representatives of this class are *a*-marked consistently.

The example *a Madonna* (map no. 1904, ALEIC) shows *a*-marking of a proper name that bears a definite article. However, nominals that bear an overt definite article usually do not get *a*-marked. Also, Marcellesi (1986) concludes that names of saints or the like usually do not bear an overt article and are consistently *a*-marked. In the corpus of Neuburger (2008), the only example of a religious figure, *Diu* 'God', occurs with a definite article and does not bear *a*-marking.

The results from the QIab survey are less clear-cut than the descriptions in the literature. The strategy of gathering the data exclusively under the categorial criterion 'proper name[human+]', obscured important differences between the two examples under investigation (for this category, test sentences nos. Iab.13a and b, Iab.14a and b, and Iab.16a and b, were considered; see the questionnaire in the appendix, Chapter 9). Essentially, *Ghjesu Cristu* and *a Madonna* have the same semantic status. The differences result from structural differences, as *a Madonna* bears a definite article. I tested Bottiglioni's *a Madonna* example in my survey, comparing it with *Ghjesu Cristu*. While the majority of informants considered *a*-marking of *Ghjesu Cristu* to be 'good', it was not so simple for the *Madonna* examples; where *Madonna* bore the definite article $a_{[DET.FEM.SG]}$ (see Table 21), *a*-marking was not the preferred construction and eight informants even excluded it as 'bad'. One informant proposed to delete the article and allow for *a*-marking only (see Iab.13c/alternative in Table 21).

(45) *Prighemu   à/ø   a   Madonna*

worship.3PL   DOM/Ø   DEF.ART.   Madonna

'We worship the Madonna'

Table 21: Ghjesu Cristu and a Madonna (QIab test sentences Iab.13a–c, and Iab.14a and b)

'We worship …'		total	***good***	*bad*	?
Iab.13a	*Prighemu à a Madonna.*	46	**9**	8	2
Iab.13b	*Prighemu ø a[DET] Madonna*	46	**37**	2	2
Iab.13c/ alternative	*Prighemu à (*a[DET]) Madonna*		**1**		
Iab.14a	*Prighemu à Ghjesu Cristu*	46	**40**	1	0
Iab.14b	*Prighemu ø Ghjesu Cristu*	45	**10**	7	2
Total			**97**		

Like Example 35, the Madonna example may be influenced by haplology.

6.2.2 *Proper nouns [human]*

In the literature, proper nouns referring to humans is the class most consistently reported to be affected by *a*-marking. The data from Neuburger (2008) confirm that finding. These claims are based on 'bare' proper nouns only, without any prenominal items.

The majority of QIab informants (40 out of 46 tokens) considered the *a*-marked example 'good' (Table 22), whereas few considered the unmarked sentence to be 'good'.

(46) *(In piazza:)*   *Cunnoscu*   *à/ø*   *Battistu*   *ma*   *un*   *cerc'à*   *ellu.*

In the square: know.1SG DOM/ø Battistu but not look.for.1SG_DOM him

'In the square: I know Battistu, but I'm not looking for him'

Table 22: Battistu [male person] (QIab test sentences Iab.17a and b)

'I know Battistu…'		total	***good***	*bad*	?
Iab.17a	*Cunnoscu à Battistu…*	46	**40**	0	0
Iab.17b	*Cunnoscu ø Battistu…*	48	**8**	11	2
Total			**48**		

If we consider human proper nouns used metonymically, such as using the name of an author to refer to her/his work, we see that the rate of *a*-marked nominals judged as 'good' diminishes slightly (only 36 out of the 46 tokens were considered 'good', see Table 23), compared to proper nouns[human] in general (40 out of 46 tokens, see Table 22).

(47) *Leghju*   *à/ø*   *Goethe*

Read.1SG DOM/ø Goethe

'I read [(a) book(s) of] Goethe'

Table 23: Goethe [metonymically used to refer to his book(s)] (QIab test sentences Iab.35a and b)

'I read Goethe'		total	***good***	*bad*	?
Iab.35a	*Leghju à Goethe*	46	**36**	2	1
Iab.35b	*Leghju ø Goethe*	45	**13**	8	0
Total			**49**		

Not all proper nouns used metonymically behave the same. While proper nouns[human] proper nouns used metonymically, as in Table (23), are *a*-marked (and ranked 'good'), toponyms used as song titles are never *a*-marked: of three examples in Neuburger (2008), none of them is *a*-marked.

6.2.3 *Proper nouns [animate]*

Proper nouns referring to animals are rarely mentioned in the literature. In most cases animals are not listed separately. Marcellesi (1986) quotes animal names and distinguishes between those bearing an overt definite article as part of their name (which are not *a*-marked), and bare animal names without an overt definite article (which, by contrast, are *a*-marked). Also, Albertini (1972: 92) mentions animal names among proper nouns which are *a*-marked. The two occurrences of animal names in Neuburger (2008) both appear with *a*-marking (cf. Neuburger & Stark 2014: 376).

The results from the questionnaire show a similar picture for proper nouns[animate] and proper nouns[human]. For test sentence number Iab.31, the majority of informants (34 out of 45 tokens) chose the *a*-marked proper nouns for animals as the best option. However, almost the half of the tokens (22 out of 45) regarding the unmarked option were also considered 'good' (see Table 24).

(48) *(Chì faci?)*   *Guardu*   *à/ø*   *Lione,*   *u*   *cane*   *di*   *Petru:*

what do.2SG   watch.1SG   DOM/ø   Lione   DEF.ART.   dog   of   Petru:

'(What are you doing?) I watch Lione, the dog of Peter:

ha   *manghjatu*   *un*   *scarpu.*

have.3SG   eat.PAST.PART   INDEF.ART.   shoe

he has eaten a shoe'

Table 24: Lione [dog name] (test sentences nos. 31a and b)

'I watch Lione, the dog…'		total	*good*	*bad*	?
Iab.31a	*Guardu à Lione, u cane…*	45	**34**	2	0
Iab.31b	*Guardu ø Lione, u cane…*	45	**22**	8	2
Total			**56**		

The picture becomes more diverse if we introduce a proper noun that includes a definite article as part of the name: *U Sciroccu*, the name of a horse from a well-known children's story, which I took from Marcellesi (1986: 137). In addition to Marcellesi's original sentence, I also offered an option where *a*-marking and the definite article co-appeared (see test sentence no. Iab.34a). Here, the 'good'-rating of *a*-marking decreased from 34 out of 45 tokens (see the bare proper[animate] noun *Lione* in Table 24, above) to only 20 out of 46 (see Table 25, below). We might have expected informants to prefer the unmarked example, but this option was even less popular (7 out of 46 tokens).

Table 25: U Sciroccu [horse name] (test sentences nos. 34a–c)

'Petru has chosen *U Sciroccu…*'		total	*good*	*bad*	?
Iab.34a	*Petru s'hè coltu à U Sciroccu*	46	**20**	4	1
Iab.34b	*Petru s'hè coltu ø U Sciroccu*	46	**7**	11	0
Iab.34c/ alternative	*Petru s'hè coltu à (*U) Sciroccu*		**9**		
Total			**36**		

It is noteworthy that 9 informants commented on the sentence, adding that the definite article *u* of *U Sciroccu* cannot be part of the name and must be deleted. As an alternative these informants proposed unanimously an *a*-marked sentence: *Petru s'hè coltu à Sciroccu…*

The idea that proper nouns bear a definite article at all seemed to be difficult for some informants. This complex state of affairs corresponds to the description in the literature where opinions on the topic diverge.

6.2.4 *Proper nouns [inanimate]*

The class of proper nouns referring to inanimate entities is a broad one. But only some of them are mentioned in the literature, such as names of months, weekdays, or years (cf. Casta 2003). These are generally considered to be *a*-marked. In the corpus of Neuburger (2008) there is only one example with a metonymically used year, *scunsiderà à 68* 'to disregard [the year] 68'

(Source: Nazione, 'Attualità', no. 14, 05/2008).

Toponyms are mentioned more often, and are considered to be *a*-marked as long as they do not bear a definite article. Nothing has been said about proper nouns that consist, for example, of brand names, such as car brands (e.g. *Fiat*) or a washing powder (e.g. *Persil*). As proper nouns are a relevant class for Corsican *a*-marking in general, I included brand names in the survey (QIab test sentences Iab.04a and b, Iab.11a–c, Iab.12a and b, Iab.18a and b, Iab.19a–f, Iab.20a and b, Iab.24a–d, 27a and b, Iab.28a and b, Iab.30a–c, Iab.32a and b, Iab.33a and b, and Iab.45a and b; see appendix, Chapter 9).

The overall picture given in Table 26 includes all of these different kinds of proper nouns[inanimate].

Table 26: Proper nouns [inanimate]

	total	***good***	*bad*	?
à + proper noun[inanimate]	731	**249**	119	24
ø + proper noun[inanimate]	867	**495**	62	39
Total		**744**		

Informants did not show a preference for *a*-marked proper nouns[inanimate]. Not only was the number of these rated 'good' rated relatively low (249 out of 731 tokens), but also the number evaluated as 'bad' was relatively high (119 out of 731 tokens). The majority of informants considered the proper nouns[inanimate] unmarked to be 'good'. Nevertheless, the distribution was not clear-cut. The pooling of different subclasses of proper nouns[inanimate] under a single category hides a great deal of variation. To get a clearer picture I discuss examples of different subclasses in Tables 27–39.

In Table 27 we can see the pattern for the month name *settembre* 'September'. In the questionnaire, *settembre* refers specifically to September 2011, given by the context (when 'we will marry'). More informants chose *a*-marking over the definite article in this context, resulting in 32 versus 7 'good' ratings. Additionally, 11 informants gave, as an alternative, a 'bare' option without any prenominal item.

(49) *Chì mese avete sceltu per a festa?*

    'Which month did you choose for the celebration?'

    Avemu   *sceltu*   *à/u*   *settembre*

    have.3PL   choose.PAST.PART DOM/DEF.ART.   September

    'We have chosen September'

Table 27: *settembre* 'September' (QIab test sentences Iab12a–c)

'We have chosen September'		total	***good***	*bad*	?
Iab.12a	*Avemu sceltu à settembre.*	46	**32**	7	2
Iab.12b	*Avemu sceltu u settembre.*	46	**7**	10	1
Iab.12c/ alternative	*Avemu sceltu ø settembre*		**11**		
Total			**50**		

Even if the examples do not pattern the exact same way, we can observe parallels between month names and weekdays. For 'Monday' too (Table 28), examples with the definite article were considered 'good' by just 12 informants, while *a*-marked and bare examples resulted in much higher numbers, each getting 25 'good' ratings.

(50) *Aspettu*   *à*   *luni,*   *ghjè*   *a*   *festa*   *di*   *Anna.*

    wait.1SG   DOM   Monday   there_is DEF.ART party   of   Anna

    'I'm waiting for Monday, when it will be Anna's party'

Table 28: *luni* 'Monday' (QIab test sentences Iab.20a–c)

'I wait for Monday…'		total	***good***	*bad*	?
Iab.20a	*Aspettu à luni…*	46	**25**	6	1
Iab.20c	*Aspettu u luni…*	46	**12**	5	2
Iab.20b	*Aspettu ø luni…*	46	**25**	1	2
			62		

For month names as well as weekdays, we may observe that the article can also be omitted. In both cases (QIab test sentences Iab.12c in Table 27, and Iab.20b in Table 28), the speaker refers to a specific date which is of personal significance to her. This is similar to how kinship terms

with a close relation to the first or second person must appear bare or with DOM, with definite articles and determiners excluded.

Among proper nouns[inanimate] referring to time, I also included the year *2011* (Table 29). Here, the majority of informants chose the option with the definite article (and no *a*-marking), i.e. 25 'good' ratings out of a total of 51 responses, while fewer informants considered either *a*-marking or the bare year to be 'good' (i.e. 13 informants in each case).

(51) *Aspiteremmu   u/à/ø   2011 per u   nostru   matrimoniu.*

  wait.1PL   DEF.ART/DOM/ø 2011 for DEF.ART. poss.1PL wedding

  'We are waiting for 2011, for our wedding'

Table 29: 2011 '[the year] 2011' (test sentences no. 11a–c)

'We are waiting for 2011…'		total	***good***	*bad*	?
Iab.11a	*Aspittaremu à 2011…*	46	**13**	8	3
Iab.11c	*Aspittaremu u 2011…*	46	**25**	1	4
	Aspittaremu ø 2011…	46	**13**	3	5
Total			**51**		

The class of brand names patterns in a very different way. While for month names and weekdays, examples with *a*-marking were ranked as 'good' by a relatively high number of informants compared to those bearing the definite article (see Tables 27 and 28), in the context of car names (Table 30), *a*-marking was not rated 'good'; in combination with a definite article, *a*-marking resulted in zero 'good' answers, while *a*-marking in combination with a demonstrative was chosen as 'good' by just one informant, and *a*-marking without other prenominal items was only considered 'good' by four informants. By contrast, options with definite articles or demonstratives only (i.e. without *a*-marking) resulted in high numbers of 'good' ratings.

Table 30: Fiat (QIab test sentences Iab.24a–d, and Iab.04a and b)

'…bought the/this new Fiat'		total	*good*	*bad*	?
Iab.24a	*…cumpratu à a*[DET] *Fiat nova.*	46	**0**	12	2
Iab.24b	*…cumpratu à issa*[DEM] *Fiat nova.*	46	**1**	11	1
Iab.24c	*…cumpratu ø a*[DET] *Fiat nova.*	46	**41**	0	2
Iab.24d	*…cumpratu ø issa*[DEM] *Fiat nova.*	46	**39**	0	0
Total			**81**		
'…damaged the new Fiat of Anna'					
Iab.4a	*…dannighjatu à Fiat nova di Anna.*	46	**4**	12	1
Iab.4b	*…dannighjatu a*[DET] *Fiat nova di Anna.*	46	**40**	1	0
Total			**44**		

This general tendency does not change if we combine car names with possessives (Table 31). Here again, the *a*-marked examples were disfavoured: where *a*-marking was combined with a definite article and the possessive, only 3 tokens were rated 'good' (QIab test sentence Iab.18a in Table 31); similarly, where *a*-marking was combined with just a possessive (i.e. without a definite article), only 6 tokens were rated 'good'. Examples without *a*-marking were chosen most often: 40 and 36 tokens, for definite article with possessive, and possessive without definite article, respectively. Affectedness in terms of verb semantics, i.e. 'to look for' versus 'to overturn', did not seem to influence the result.

Table 31: mo Peugeot (QIab test sentences Iab.18a and b, and Iab.27a and b)

'looking for my Peugeot…'		total	*good*	*bad*	?
Iab.18a	*…cercu à a*[DET] *mo Peugeot…*	46	**3**	9	2
Iab.18b	*…cercu a*[DET] *mo Peugeot…*	46	**40**	1	2
Total			**43**		
'…overturn my Peugeot'					
Iab.27a	*…lampatu à mo Peugeot.*	45	**6**	11	1
Iab.27b	*…lampatu a mo Peugeot.*	45	**36**	2	2
Total			**42**		

One can see that also other brand names, such as the washing powder *Persil* (Table 32) pattern similar to car names, and differently to month names or weekdays. Here again, the example with the definite article was favoured (28 'good' rated tokens).

(52) *In magazinu:*

'In the store:'

Aghju   cumpratu   à/u/ø   Persil   hè in offerta è u più ecunomicu detergente chi ci hè

have.1SG   buy.PAST.PART   DOM/DEF.ART./ø Persil

'I have bought Persil, it is on offer and the cheapest washing powder there is [in the store]'

Table 32: Persil [name of washing powder] (test sentence no. 30a–c)

'I have bought Persil…'		total	***good***	*bad*	?
Iab.30a	...*cumpratu à Persil*...	46	**12**	10	1
Iab.30b	...*cumpratu u*[DET] *Persil*...	46	**28**	3	1
Iab.30c	...*cumpratu ø Persil*...	46	**12**	9	2
Total			**52**		

Toponyms behave differently. According to the literature toponyms may be *a*-marked or not: as mentioned before regarding proper nouns, toponyms too can only be *a*-marked if they do not bear an overt definite article (as part of the name). In Neuburger (2008) out of 13 occurrences 11 are *a*-marked; one example bears neither *a*-marking nor definite article, the other one is metonymically as a song title (see remarks on metonymy, section 6.2.2).

The data from QIab show that the informants valued unmarked toponyms as 'good' (111 tokens) almost as often as *a*-marked ones (126 tokens) (see Table 33).

Table 33: toponyms (QIab test sentences Iab.19a–f , Iab.32a and b, and Iab.45a and b)

	total	***good***	*bad*	?
à + toponym	226	**126**	22	8
ø + toponym	225	**111**	23	11
Total		**237**		

Not included in the overview in Table 33 is one toponym, in which the definite article *(l)a* is considered a constituent of the toponym, i.e. *A Marana*, a small confined area on the north-east coast of Corsica. The results of this example pattern in a very different way (Table 34). The overwhelming majority of informants preferred the unmarked example, comprising 41 of the 45 'good'-rated tokens.

169

Table 34: (L)A[71] Marana (QIab test sentences Iab.33a and b)

'The smoke covers *A Marana*'		total	***good***	*bad*	?
Iab.33a	*...copre à La*[DET] *Marana.*	46	**4**	6	2
Iab.33a	*...copre A*[DET] *Marana.*	46	**41**	2	1
Total			**45**		

In contrast to the horse name U Sciroccu (see Table 25), or to A Madonna (Table 21), none of the informants here proposed that the article should be deleted as an alternative to *a*-marking.

6.2.5 *Common nouns [human]*

In the literature, the relevant feature with regard to common nouns and DOM is the syntactic structure of the nominal: the presence of a definite article, which is incompatible with the *a*-marker. Common nouns as a nominal class are excluded from *a*-marking, the reason being that they co-occur with a definite article. This is described by various authors (see sections 4.1 and 4.3). Common nouns that diverge from this structure (e.g. bare nouns) and the way these interact with the *a*-marker have not been considered at all. In the corpus of Neuburger (2008) Igathered 74 bare nouns in DO-function, 16 referring to humans, 5 to animates and 53 to inanimates. Of these, only one bare common noun[human] is *a*-marked (see example (53)).

(53) […] *ùn mette   à   moglie incinta*   (*Mitulugia*: 39)

    not make.3SG   DOM   wife pregnant

  '[…] not to get his wife pregnant'

In the context to the examples, the noun *moglie* 'wife' refers to a specific person, closely related to the protagonist in the text; she is his wife. *Moglie* seems to be used similarly to kinship terms and thus gets *a*-marked (see section 6.1 on kinship terms, above). This is not the case for any other bare noun[human] in Neuburger (2008).

    The same holds for common nouns. We saw in the earlier sections on kinship terms and proper nouns (sections 6.1 and 6.2, respectively) that examples which show both the *a*-marker and the definite article were rated as 'good' less often then examples with the *a*-marker only.

    But prenominal items, such as definite articles and demonstratives, provoke a notably high number of 'bad' evaluations (22 and 9 tokens for definite articles and demonstratives,

---

[71] The definite article feminine singular may change here from *a* to *la* due to phonological reasons, i.e. as it is preceeded by the *a*-marker.

respectively).

Table 35: Common nouns [human]

	total	***good***	*bad*	*?*
à + def.art.	90	**11**	22	1
ø + def.art.	45	**39**	0	0
à + demonstrative	41	**11**	9	0
ø + demonstrative	41	**28**	2	2
à + negative quantifier	45	**15**	7	1
ø + negative quantifier	45	**18**	5	1
Total		**122**		

For the QIab survey, for each of the combinations in Table 35 above, I gave one example, following their order of appearance in the table (QIab test sentences Iab.42a and b, Iab.43a and b, and Iab.44a and b):

		total	***good***	*bad*	*?*
Iab.42b	*E cundizioni ughjinche attiranu à i spiculatori*	45	**6**	10	1
Iab.42a	*E cundizioni ughjinche attiranu i spiculatori*	45	**39**	0	0
Iab.44a	*Cunnoscu à iss'omu*	41	**11**	9	0
Iab.44b	*Cunnoscu iss'omu*	41	**28**	2	2
Iab.43a	*Ùn vegu à nisciun'omu*	45	**15**	7	1
Iab.43b	*Ùn vegu nisciun'omu*	45	**18**	5	1

I also tested a negative quantifier in prenominal position, (QIab test sentence Iab.43). The example was inspired by Marcellesi (1986: 138, ex. 33; see also appendix, Table 60). If we compare the results for the first two classes (i.e. definite articles and demonstratives) with those for prenominal negative quantifiers, we see that the picture changes. The results for prenominal negative quantifiers with or without *a*-marking diverged less than those for definite articles and demonstratives. It may be that *nisunu* 'nobody' triggered *a*-marking irrespective of its prenominal or pronominal function, because the informants did not consider it a determiner in prenominal position that would prevent *a*-marking (compare with negative quantifier as a proform, section 6.4.6, below).

6.3 Proforms

In the literature, the only mentions of the class of proforms (i.e. nominals items that are not full lexical nouns, nor proper nouns or kinship terms, but are linked to another nominal introduced at some point in the discourse) are of personal pronouns and negative quantifiers such as *nimu* 'nobody' and *nulla* 'nothing'. Of the known DOM-triggers, only humanness of the nominal's referent is reported as relevant for proforms. The literature makes no mention of definiteness, nor about the syntactic structure of the proforms.

For most full lexical nouns, the animacy value is part of the lexical or inherent information of the DO-nominal. This holds also true for most proper nouns (disregarding metonymic uses, such as toponyms used for groups of persons, or the like). For proforms, we may distinguish two kinds with regard to animacy. Some proforms are used exclusively to refer to humans, such as personal pronouns. Others have two 'sets' of lexical items. One set is specialised for human referents, such as the interrogative *quale* 'who', the universal quantifier *ognunu,a* 'everybody', the indefinite quantifiers *qualchidunu* 'someone' and *qualchissia* 'who(m)ever', and the negative quantifiers *nimu, nisunu* 'nobody' – for these items, humanness is part of the lexical information or an inherent property. For others this does not hold true – the universal quantifier *tutti* 'all, everybody', the demonstratives *quistu/quissu* 'this' and *quillu/quiddu* 'that', and the indefinite *unu* 'one', may also refer to (in)animates.

In the sections that follow, I take as a point of reference the notion of a referentiality hierarchy (see section 2.1), starting with the leftmost class of items, i.e. the most specific, most definite and human items: personal pronouns. I proceed from there to items with a decreasing degree of definiteness and/or specificity and/or humanness, before concluding with negative quantifiers.

6.3.1 *Personal pronouns*

In the literature, personal pronouns are among the few explicitly mentioned proforms for which DOM is relevant; they are consistently *a*-marked in DO-function. This was also confirmed in Neuburger (2008), wherein all personal pronouns are *a*-marked (see also Table 11, and Neuburger & Stark 2014: 376).

The data from QIab confirm these findings: the vast majority of informants considered the *a*-marked example 'good' (40 tokens) compared to the unmarked example which resulted in eight 'good' tokens (see Table 36). There were also a remarkably high number of '?'-judgements: 28 informants were not sure about the unmarked personal pronoun.

(54) Iab.17ab *Cunnoscu à Battistu ma*

ùn   cercu   à/ø   ellu

not   looking.for.1SG   DOM/ø   him.PERS.PRON.3MASK.SG

'I know Battistu, but I'm not looking for him'

Table 36: Personal pronouns (QIab test sentences Iab.17a and b)

	total	**good**	*bad*	?
à + personal pronoun	46	**40**	0	6
ø + personal pronoun	48	**8**	11	28
Total		**48**		

6.3.2 *Demonstratives [human]*

Unlike personal pronouns, which are exclusively used to refer to humans, demonstratives usually refer only to animates or inanimates. When used for human referents, they always bear always a certain, mostly negative, connotation. This, too, is mentioned by Marcellesi (1986: 133); for more detail, see section 4.2 on DOM and humanness. He considers demonstratives a domain of optional marking; in his list of examples (cf. Marcellesi 1986: 138) he gives an *a*-marked demonstrative proform.

Of the examples in Neuburger (2008), none of the demonstrative proforms get *a*-marked, irrespective of animacy; 7 demonstratives[human] and 9 demonstratives[inanimate] are all unmarked. The 7 demonstratives[human] are followed by a restrictive relative clause; any (cor)relations between the non-marking and the relative clauses have not been further analysed.

In the QIab questionnaire, I included a demonstrative[human] proform, shown in example (55).

173

(55) *Dui omi anu arrubatu i soldi di a signora Muralli. À qual'aghju da seguità?*

'Two men have pinched money from Signora Muralli. Which one do I have to follow?"

Adhju	*da*	*seguità*	*à/ø*	*quistu*	*o*	*à/ø*	*quillu?*
have.1SG	to	follow	DOM/ø	this	or	DOM/ø	that?

'Do I have to follow this one or that one?'

The results are given in Table 37. Most informants (37 tokens) chose the *a*-marked example as 'good',. Only six informants considered the unmarked option 'good'.

Table 37: Demonstratives *quistu, quillu* [human] (QIab test sentences Iab.06a and b)

'Have I to follow this or that one?'		total	***good***	*bad*	?
Iab.6a	*Adhju da seguità à quistu o à quillu?*	45	**37**	2	0
Iab.6b	*Adhju da seguità ø quistu o ø quillu?*	45	**6**	7	6
Total			**43**		

Compared to the distribution patterns seen in previous categories, this result shows a strong preference for *a*-marking. Demonstrative proforms do not seem a domain of optional *a*-marking.

6.3.3 *Relative pronoun [human]*

According to the literature, the relative pronoun *chi/chì* 'whom, what' can refer to any kind of referent irrespective of its animacy value. Like other proforms, the relative proform *chi/chì* is also seldom mentioned in the literature. In fact, only Marcellesi (1986: 133–134) explores the topic in any detail, as *chi/chì* poses some problems according to his analysis: compared to other proforms, where humanness is most relevant for *a*-marking, the relative proform *chi/chì* does not get *a*-marked at all, regardless of the animacy value of its referent.

The two occurrences of *chi/chì* in the corpus of Neuburger (2008) both refer to humans and are both *a*-marked. I used one of these, in a slightly modified form, in the QIab questionnaire. The result of this was very similar to that for demonstratives[human]: the majority of informants rated the *a*-marked example as 'good' (37 tokens).

174

Table 38: Relative pronoun [human]: *chi/chì* (QIab test sentences Iab.36a and b)

'I give money to the poor, to help whomever needs it'		total	***good***	*bad*	?
Iab.36a	*Dò i soldi à i poveri, per aiutà à chì ne ha bisognu.*	44	37	3	0
Iab.36b	*Dò i soldi à i poveri, per aiutà ø chì ne ha bisognu.*	44	8	6	1
Total			**45**		

Thus, Marcellesi's (1986) statement is not confirmed by the data from this questionnaire: in fact, the proportion of respondents, rating *a*-marked relative pronouns[human] as 'good', is high (37 out of 44 tokens).

6.3.4 *Interrogatives*

The proform *chi/chì* is not only a relative pronoun, as seen in the previous section; a homophonous form *chi/chì* is also found among the interrogatives.

According to Marcellesi (1986), optional *a*-marking in Corsican is only found with proforms. He cites pronominal demonstratives referring to humans (see section 6.3.2), and the interrogative *chi/chì*. Descriptions of the latter are rather imprecise, not only with regard to DOM (and not only with regard to the description by Marcellesi 1986), but in general. Does it refer to humans exclusively or also to animates and inanimates? Does it appear in specific syntactic functions or is it multifunctional, i.e. can it be subject and/or DO?

To clarify the use of *chi/chì*, I included some examples in the questionnaire. I compared *chi/chì, quale* and *à quale* in the otherwise identical question *Chì/quale/à quale hè ch'hè vinutu oghje à u scagnu? He vinutu u sgiò Baldoni* 'Who came to the office today? It was Mr Baldoni'. I wanted to know (i) which item is preferred for [human] referents and (ii) if the informants distinguish items according their function, i.e. between subject and DO items.

Comparing *quale* and *chi/chì* for subject function, *quale* was the favoured lexical item: 44 informants rated *quale* as 'good', while only 2 informants considered *chi/chì* a 'good' option, and just 1 informant chose *à quale* (i.e. the *a*-marked *quale*) as the best option (see Table 39).

Table 39: *Interrogatives[subject]* à quale – chì – quale *(QIab test sentences Iab.08a–c)*

'Who came to the office today?'		total	***good***	*bad*	?
Iab.8a	*À quale hè ch'hè vinutu oghje à u scagnu?*	46	**1**	13	0
Iab.8b	*Chì hè ch'hè vinutu oghje à u scagnu?*	46	**2**	11	1
Iab.8c	*Quale hè ch'hè vinutu oghje à u scagnu?*	46	**44**	1	0
Total			**47**		

With regard to *quale* in DO-function, the sentence with the *a*-marking are predominantly considered 'good': see Table (40).

Table 40: à/ø quale [human] in DO-function (QIab test sentences Iab.6a and b)

'Whom do I have to follow?'		total	***good***	*bad*	?
Iab.6a	*À qual'aghju da seguità?*	45	**27**	2	0
Iab.6b	*Qual'aghju da seguità?*	45	**6**	7	3
Total			**33**		

6.3.5 *Universal quantifiers [human]*

Universal quantifiers correspond to different lexical items, some of which are inherently or lexically specified for humanness, such as *ognunu,a* 'everybody', while others are underspecified for animacy, such as *tutti* 'all'. In the literature, humanness of the DO referent is also considered most relevant for *a*-marking of universal quantifiers. This was confirmed by the data from Neuburger (2008), wherein all universal quantifiers[human], including *tutti* and *ognunu,a*, were *a*-marked (5 out of 5).

For this study, I tested two different universal quantifiers for Corsican. *Ognunu,a* 'everybody', which is exclusively used for [human] referents, bears the value [human] inherently. By contrast, *tutti* 'all' can refer to any kind of plural entity, whatever the animacy value of the referents.

The general picture (Table 41) showed a strong preference for *a*-marked universal quantifiers; 99 *à*-marked tokens were rated 'good', compared to the unmarked universal quantifier, which was considered 'good' by 26 informants.

176

Table 41: Universal quantifiers [human] (QIab test sentences Iab.37a and b, Iab.39a and b, and Iab.50a and b)

	total	*good*	*bad*	?
à + universal quantifier	120	**99**	4	3
ø + universal quantifier	120	**26**	18	4
Total		**125**		

If we look at *tutti* and *ognunu,a* in isolation, we see two divergent patterns for the two lexical items. For *ognunu,a* 'everybody', a slight majority of informants considered *a*-marking 'good' (29 tokens) compared the unmarked option (20 tokens); see Table 42.

(56) *Chjamu   à/ø   ognunu   à   tene   u   rollu chì   hè   u   soiu*

invite.1SG   DOM/ø   everybody   to   take   DEF.ART part   that be.3SG   DEF.ART poss.3SG

'I invite everybody to take the part that is hers/his'

Table 42: ognunu,a [human] (test sentences nos. 39a and b)

'I invite everybody…'		total	*good*	*bad*	?
Iab.39a	*Chjamu à ognunu …*	45	**29**	3	2
Iab.39b	*Chjamu ø ognunu …*	45	**20**	5	1
Total			**49**		

By contrast, the distribution of answers is much more clear-cut for *tutti*[human]: the vast majority of informants considered the *a*-marked example 'good' (70 tokens out of a total of 76 'good' tokens), while the unmarked examples were rated 'good' by just 6 informants. A substantial number (13 informants) evaluated the unmarked option as 'bad':

Table 43: tutti [human] (QIab test sentences Iab.37a and b, and Iab.50a and b)

	total	*good*	*bad*	?
à tutti	75	**70**	1	1
ø tutti	75	**6**	13	3
Total		**76**		

6.3.6 *Indefinites*

Indefinite proforms have been well documented in the literature. Most authors cite data with

qualchissia 'someone', *qualchidunu* 'somebody'. Both items are inherently or lexically specified for humanness. Other indefinite proforms, such as *unu* 'one', which are underspecified for animacy and therefore able to refer to any entity, are not mentioned in the literature.

In the results of Neuburger (2008), all three occurrences of [human] *unu* are not *a*-marked, irrespective of specific references or generic interpretation. Also, the [animate] *unu,* referring to a dog specified by the context, does not get *a*-marked. No instances of *qualchidunu* 'somebody' or *qualchissia* 'someone' were found in the corpus of Neuburger (2008).

In the QIab questionnaire, sentences were presented with different lexical items: on the one hand, I tested lexical items which exclusively refer to humans, i.e. *qualchissia* 'anybody, somebody' and *qualcunu, qualchidunu* 'somebody, someone'; on the other hand I tested *unu* '(some)one', which may refer to human and (in)animate entities alike.

Almost three-quarters (232) of all (305) 'good'-valued tokens were accounted for by *a*-marked options, and almost one quarter (73 tokens) were accounted for by the unmarked option. However, of the 312 tokens relating to unmarked indefinites, 43 were also considered 'bad', see Table (44).

Table 44: Indefinites [human] (QIab test sentences Iab.02a and b, Iab.03a and b, Iab.05a–e, Iab.15a and b, and Iab.23a–g)

	total	*good*	*bad*	?
à + indefinite	314	**232**	13	14
ø + indefinite	312	**73**	43	20
Total		**305**		

Looking at the different subclasses of indefinites lexically specified for humanness (i.e. *qualchissia, qualchidunu, qualcunu*) and underspecified for humanness (*unu),* in both cases the *a*-marked structure was the preferred pattern.

For the former, more than three-quarters (174) of the 223 'good' ratings were assigned to the *a*-marked option (see Table 45). Nevertheless, a substantial number of informants also considered unmarked indefinites referring to humans to be 'good' (49 tokens out of 223 tokens):

Table 45: Indefinites lexically [human]: *qualchissia, qualchidunu, qualcunu*

	total	***good***	*bad*	?
à + indefinite	224	**174**	6	9
ø + indefinite	223	**49**	30	16
Total		**223**		

As for the latter, two-thirds of the informants rated *a*-marked *unu* 'good' (see Table 46). One quarter considered the unmarked item 'good'.

Table 46: unu [human]

	total	***good***	*bad*	?
à unu	90	**58**	7	5
ø *unu*	89	**24**	13	4
alternative *ne* partitive + ø *unu/*	46	**5**		
Total		**87**		

Remarkably, 5 out of 46 informants offered an alternative construction: they gave a partitive construction with the particle *ne,* and combined it with the unmarked *unu*: […] *ne cunnosci ø unu chì* […], 'You know *ne*PART one who […]'. The same structure was proposed by a few informants for *unu* referring to animates (see Table 48, below).

With *unu* I also tested whether verb mood, i.e. the distinction between indicative and subjunctive mood, had any impact on the acceptability of *a*-marking, as has been shown in other languages with DOM-systems, such as Spanish.

(57) *Cunnosci   à/ø   unu   chì   sà/sappii   riparà   a   mo   vittura?*

    know.2SG   DOM/ø   one   who   know.IND/know.SUBJ   repair   DEF.ART.   POSS   car

    'Do you know someone who knows how to repair my car?'

In the literature, verb mood is not mentioned as a relevant factor for *a*-marking in Corsican. This was confirmed by the results from the QIab study, as seen in Table 47:

Table 47: Verb mood: subjunctive versus indicative (test sentences nos. 5a–e)

'…who knows how to repair my car?'		total	***good***	*bad*	?
Iab.5a	*Cunnosci à unu chì sappii*[SUBJ] *riparà…*	45	**31**	3	3
Iab.5b	*Cunnosci ø unu chì sappii*[SUBJ] *riparà…*	44	**11**	6	2
Iab.5d	*Cunnosci à unu chì sà*[IND] *riparà…*	45	**27**	4	2
Iab.5e	*Cunnosci ø unu chì sà*[IND] *riparà…*	45	**13**	7	2
Total			**82**		

Informants were asked to choose one sentence for each mood, i.e. either the *a*-marked or unmarked sentence with the verb in subjunctive mood (QIab test sentences Iab.05a or b), and likewise with the verb in indicative mood (QIab test sentences Iab.05d and e). Comparing the numbers of *a*-marked examples rated 'good', there was little difference between subjunctive mood (31 tokens) and indicative mood (27 tokens). The unmarked examples which were considered 'good' also show no marked difference between subjunctive mood (11 tokens) and indicative mood (13 tokens).

I also tested *a*-marking of *unu* with an [animate] referent, i.e. a dog. The context introduces two specific dogs, one with a possessive pronoun, related to the 1st person, and a second one modified by a possessive PP; see example (58) and the accompanying Table (48):

(58) *U mo cane ghjè nant'à strada, si batte cù u cane di u vagabondu.*

'My dog is on the street and is fighting with the dog of the vagrant'

Petru   hà   tumbatu   à   unu   senza   ferisce   l'altru.

Petru   have.3SG   strike.PAST.PART   DOM/ø   one   without   injure   DEF.ART_other

'Petru has struck one without injuring the other'

Table 48: unu [animate] (dog) (QIab test sentences Iab.40a–c)

'Petru has struck one…		total	***good***	*bad*	?
Iab.40a	*Petru hà tumbatu à unu …*	45	**13**	6	1
Iab.40b	*Petru hà tumbatu ø unu …*	45	**18**	3	2
Iab.40c/ alternative	*Petru ne hà tumbatu ø unu …*		**11**		
Total			**42**		

More informants (18 tokens) rated the unmarked option as 'good' than the *a*-marked option

(13 tokens). As seen earlier, for this example too, informants offered an alternative with a partitive construction and without *a*-marking of *unu*, in this case representing a quarter of the responses (11 tokens), which is remarkable. In both Tables 46 and 48, *unu* refers to a subset (of someone able to repair the car; of dogs) and may trigger the partitive construction with *ne*.

6.3.7 *Negative quantifiers*

Within the class of negative quantifiers we have three lexical items that differ with respect to the animacy value of the referent; *nimu* and *nisunu,a* 'nobody' refer to humans exclusively (Table 50), while *nulla* 'nothing' is applied for inanimates (Table 51). Negative quantifiers, or rather the lexical items *nimu* 'nobody' and *nulla* 'nothing', are frequently mentioned in the literature. While for *nimu a*-marking is considered mandatory, *nulla* is excluded from *a*-marking.

This description of the distribution of *a*-marking is in accordance with the findings of Neuburger (2008), wherein all occurrences of *nimu* and *nisunu* 'nobody' are consistently *a*-marked, while *nulla* 'nothing' is never *a*-marked.

The data from QIab show that *a*-marking of [human] negative quantifiers resulted in a large number of 'good'-judgements, corresponding to 78 tokens, and only 17 'good' ratings for the unmarked version. Table 49 shows the general overview of negative quantifiers:

Table 49: Negative quantifiers [human] (QIab test sentences Iab.21a and b, and Iab.22a and b)

	total	***good***	*bad*	?
à + negative quantifier	91	**78**	4	1
ø + negative quantifier	91	**17**	13	3
Total		**95**		

If we distinguish between the two lexical items representing this class of proforms, *nimu* and *nisunu* (see Table 50), there was a strong preference for *a*-marking of both items. Rating *à nimu* as 'good' was selected by the highest proportion of total informants, compared to all other items under investigation (i.e. 43 out of 46 tokens, equivalent to 93.48%), while *à nisunu* was rated as 'good' by 35 of the 46 informants (equivalent to 77.78%).

Table 50: nimu versus *nisunu* (QIab test sentences Iab.21a and b, and Iab.22a and b)

'I see nobody'		total	***good***	*bad*	?
Iab.21a	*Ùn vecu à nimu.*	46	**43**	2	0
Iab.21b	*Ùn vecu ø nimu.*	46	**6**	8	1
Iab.22a	*Ùn vecu à nisunu.*	45	**35**	2	1
Iab.22b	*Ùn vecu ø nisunu.*	45	**11**	5	2
Total			**85**		

Despite the differences between the two items *nimu* and *nisunu*, the preference for *a*-marking is quite clear for negative quantifiers referring to humans. We also see a distinct pattern for negative quantifiers referring to inanimates. But the results show a reversed distribution (Table 51): the majority of informants (41 tokens) considered unmarked *nulla* 'good', and only a very few (i.e. 4 informants) judged *a*-marked *nulla* 'good'.

Table 51: nulla [inanimate] (QIab test sentences Iab.25a and b)

'I see nothing'		total	***good***	*bad*	?
Iab.25a	*Ùn vecu à nulla.*	46	**4**	14	1
Iab.25b	*Ùn vecu ø nulla.*	46	**41**	0	0
Total			**45**		

6.4 Summary of QIab results

My fieldwork was motivated by two goals, which were elaborated and presented in section 4.8, and are repeated here for convenience:

1. *Quantity and quality of sentences*:
   a) to increase the number of sample sentences of the kind already presented in the literature, particularly in cases where descriptions are based on isolated examples
   b) to include new sample sentences containing semantic features and syntactic structures that are currently under- or not represented in the literature
2. *Corsican DOM in the literature*:
   - to use the new data to test existing descriptions of the phenomenon of DOM in Corsican.

With regard to the first of these issues, I increased the number of examples for all noun classes and considered different animacy values, both to test the examples from the literature and to include underrepresented structures and features. To do so, I asked informants to choose the

best option from the examples provided. This forced choice task resulted in a total of 2410 tokens corresponding to samples chosen as 'good' by informants (recorded in the tables in sections 6.1–6.4, above). These tokens were split between different word classes, animacy values and syntactic structures.

With regard to the second issue, Corsican DOM in the literature, I discussed the results in detail in the previous sections of this chapter.

An overview of the distribution of results from the forced choice task is given in Table 52, as follows:

Table 52: Overview of the forced choice task, i.e. the options chosen as 'good' by the informants

	kinship terms					proper nouns											
	à	*ø*	*à +* poss	*ø +* poss	*ø +* def.art + poss	*à*	*ø*	*à +* def.art	*ø +* def.art	*à*	*ø*	*à*	*ø*	*à*	*ø*	*à* + def.art	*ø* + def.art
[human+]						saints *total tokens: 97*											
						41	10	9	37								
[human]	human *total tokens: 267*									human *total tokens: 48*		human metonymic use *total tokens: 49*					
	68	16	84	21	78					40	8	36	13				
[animate]														animal *total tokens: 92*			
														43	22	20	7

	proper nouns																	
	à	*ø*	def. art	*à*	*ø*	def. art	*à*	*ø*	def. art	*à*	*ø*	def. art	dem	*à +* def.art	*à* + dem	*à+* def.art + poss	*ø +* def.art + poss	*à +* poss
[inanimate]	month name *total tokens 50*			weekday *total tokens: 62*			year (*2011*) *total tokens: 51*			brand names of cars *total tokens 210*								
	32	11	7	25	25	12	13	25	13	4	0	81	39	0	1	3	76	6

Table 53 (cont'd): Overview of the forced choice task, i.e. the options chosen as 'good' by the informants

	proper nouns						
	à	*ø*	**def.art**	*à*	*ø*	**def.art**	*à* + **def.art**
[inanimate]	brand names of washing powder			toponyms			
	total tokens 52			*total token 330*			
	12	12	28	149	136	41	4

	common nouns					
[human]	*total tokens: 122*					
	à + **def.art**	*ø* + **def.art**	*à* + **dem**	*ø* + **dem**	*à* + **neg.quant**	*ø* + **neg.quant**
	11	39	11	28	15	18

	proforms														
	à	*ø*	*à*	*ø*	*à*	*ø*	*à*	*ø*	*à*	*ø*	*à*	*ø*	*ø + ne*	*à*	*ø*
	personal pronouns		demonstratives		relative *chì*		interrogative *quale*		universal quantifiers		indefinites			negative quantifier	
[human]	*total tokens: 48*		*total tokens: 43*		*total tokens: 45*		*total tokens: 125*		*total tokens: 125*		*total tokens: 305*			*total tokens: 95*	
	40	8	37	6	37	8	115	6	99	26	232	73		78	17
[inanimate]											*total tokens: 42*			*total tokens 45*	
											13	18	11	4	41

185

The data from the QIab survey do not pattern as straightforwardly as the literature, and to a certain extent the data from Neuburger (2008), suggest. It was rare that all informants agreed on which alternative was best. Thus, we can only rarely conclude that acceptability is categorial. Among the few noun classes and/or examples which scored a zero on any of the judgement values, there was no single noun class which scored zero 'good' or zero 'bad' judgements; such clear preferences were found only for single exponents of classes. For example, no informant considered *a*-marking of Battistu, a proper noun[human], to be 'bad' (Table 22). Thus, no class and almost no single examples provoked a clear-cut answer among all informants.

I summarise the results in the following paragraphs, with reference to the relevant tables.

Bare kinship terms, bare proper nouns[human+] (i.e. names of religious figures), and proper nouns[human] marked by *à*, were frequently rated as 'good'; this confirms the information from the literature and from Neuburger (2008) (Tables 18–22). Prenominal items, such as possessives or definite articles, diminished the acceptability of *a*-marking of non-bare exponents, as shown by the remarkably high number of 'good' ratings for options without *a*-marking. This result also confirms traditional descriptions of Corsican DOM with regard to proper nouns and/or kinship terms. For metonymically used proper nouns[human], *a*-marking was widely accepted, although the percentage of *a*-marked options chosen as 'good' was lower than non-methonymically used proper nouns[human] (Table 23).[72]

Proper nouns[animate] also confirm these marking strategies: names of animals were preferred with *a*-marking, as long as they did not bear a definite article (Table 24). The simultaneous appearance of the definite article and the *a*-marker in the context of proper nouns[animate] seemed to cause irritation, in which cases informants preferred the deletion of the article.

The class of proper nouns[inanimate] included a broad field of subclasses, with examples of toponyms, names of years, months and weekdays, brand names of cars, and a washing powder. Only toponyms have been mentioned in the literature on *a*-marking. To the best of my knowledge, brand names have not been tested before. In my survey, acceptance for *a*-marking regarding proper nouns[inanimate] was quite low (Table 26), but the subclasses diverged significantly. On the one hand, *a*-marking of month names was widely accepted. On the other hand, 11 out of 46 informants gave a bare month name as an alternative: *Avemu sceltu ø settembre* 'we have chosen September [2011 for our wedding]' (Table 27). For weekdays, the example with *luni* 'Monday' showed an equally high percentage of tokens rated 'good' for both *a*-marked and bare *luni*. Nevertheless, we can observe a

---

72   Recall that not all metonymically used proper nouns are equally accepted with a-marking. Toponymes (as a subclass of proper nouns[inanimate]) used as song titles did not appear with a-marking.

difference: more informants evaluated the *a*-marked *luni* and considered it 'bad' (Table 28). With the year 2011, however, we see the reverse pattern (Table 29). Here, the example with the definite article was the preferred one, while *à 2011* and *ø 2011* scored comparatively low numbers of 'good' ratings. Thus, these items did not pattern consistently with the descriptions in the literature.

In the evaluation of brand names, in particular car names, informants showed a strong preference for the definite article instead of the *a*-marker; also, as for other proper nouns, the simultaneous appearance of a definite article or demonstrative with the *a*-marker was strongly disfavoured. This preference was not changed by the addition of a possessive (Table 31); also with the possessive, the simultaneous appearance of a definite article and the *a*-marker was disfavoured, as was the possessive with *a*-marking (but without a definite article), although this did show a slight increase in the rate of acceptance. Also, for the washing powder (Table 32), informants preferred the example with the definite article over the *a*-marked or the bare example.

Toponyms, as often stated in the literature, should be *a*-marked, with the well-known exception of toponyms with a definite article, for which no *a*-marking occurs. In Neuburger (2008) this holds true for most toponyms, save two occurrences that appear without *a*-marking nor a definite article.

In the survey, however, informants did not evaluate toponyms in such a straightforward way. In fact, their preferences were almost equally split between *a*-marked toponyms and unmarked toponyms (Table 33). The handling of the example *A Marana* (Table 34), a toponym with a definite article, was remarkable. We can compare *A Marana* to the examples with the proper noun[human+] *A Madonna*, and the proper noun[animate] *U Sciroccu*. For the latter two, a very large majority preferred the unmarked example (ø *A Marana*), thereby confirming the supposed preference for avoiding *a*-marking together with the definite article. But, unlike with *A Madonna* and *U Sciroccu*, none of the informants suggested removing the definite article for the *A Marana* example. Hence, while definite articles caused irritation if appearing with names of saints, humans and animals (i.e. animacy values [human+], [human] and [animate]), the definite article of *a Marana* was not deleted by any of the informants. This may confirm Durand's (2003) remarks on toponyms: that certain Corsican toponyms may bear definite articles as an integral component (see also section 4.1).

According to the literature, common nouns are excluded from *a*-marking because they are preceded by a definite article (i.e. the same rule as for proper nouns). But no further information is provided, either about the impact of other prenominal items (except in Marcellesi (1986)), or about bare common nouns; nor do we find remarks about the relevance of features, such as animacy. In the corpus of Neuburger (2008), no *a*-marking of common nouns can be found, irrespective of the animacy value of the referents. This holds true both for exponents with articles and for bare nouns

(with the exception of *à moglie* 'wife', example (53)).

The preferences of informants in my survey were less straightforward. I tested common nouns[human] with different prenominal items – i.e. definite articles, demonstratives and negative quantifiers – each presented with and without *a*-marking. In all cases there was a preference for the unmarked common noun with any of these prenominal elements. But the judgment of *a*-marking of these nominals differed considerably (see Table 35): while there were 22 tokens representing 'bad' ratings for the combination '*à* + definite article', the combination '*à* + demonstrative' was rated 'bad' by 9 informants, and the combination '*à* + negative quantifier' by 7 informants. In addition, the last option was considered 'good' by a third of all informants.

The final noun class consists of proforms. In general, in the literature, humanness is the only trigger for *a*-marking mentioned as relevant to this class. This is roughly confirmed both by Neuburger (2008) and the data from the QIab survey, with certain notable exceptions, such as variations in preferences for different lexical items belonging to the same noun class.

Personal pronouns in the QIab survey (Table 36) were in line with the literature and Neuburger (2008). Demonstratives (Table 37) were considered an optional domain by Marcellesi (1986), i.e. we should expect *a*-marked as well as unmarked exponents. In Neuburger (2008) no demonstrative[human] bears *a*-marking. However, there was a preference for the *a*-marked option within the QIab dataset. The relative pronoun *chi/chì* – which can refer independently to [inanimate], [animate] and [human] nominals alike – is mentioned only once in the literature, by Marcellesi (1986). He excludes *a*-marking for *chi/chì* independently from any animacy value. Both the data from Neuburger (2008) and the data from the QIab survey show very different results to this: both occurrences in Neuburger (2008) are *a*-marked; among the informants of QIab the *a*-marked *chì*[human] was strongly favoured (Table 38).

The information about interrogatives in the literature is sketchy: on the one hand it is not clear which lexical item(s) is/are used to refer to humans at all; on the other hand, it remains unclear which one (i.e. *quale*, *chi/chì*) is used for which syntactic functions, i.e. the subject or the DO. The data gathered for this study tell us that *chì*[relative] might refer to humans and (in)animates alike (see above, section 6.3.3). For interrogatives, *quale* was preferred over *chì*[interrogative] for human referents. Where *quale* appeared in DO-functions, the *a*-marked exponent was strongly favoured (Table 40).

In the literature, universal quantifiers are generally *a*-marked if they refer to humans, irrespective of the lexical items used, i.e. *tutti* or *ognunu,a*. The difference, for these two items, lies in how they specify with regard to animacy: the latter, *ognunu,a*, is lexically or inherently specified for humanness; *tutti*, on the other hand, can refer to any kind of noun phrase regardless of its animacy

value. Furthermore in the QIab survey, *a*-marked exponents were preferred over unmarked ones. But the two different lexical items result in quite different percentages: for *ognunu,a* (Table 42), a non-negligible proportion of informants (44.44%) rated the unmarked example as 'good', for *à tutti* 'all', the preferencewas much more straightforward with an exceptionally high number of informants rating it 'good (equivalent to 93.33%; Table 43).

Among the indefinite proforms, both lexically and inherently specified nominals referring to humans, i.e. *qualchidunu, qualcunu* 'someone, somebody', *qualchissia* 'anybody', and one other item, *unu* 'one', are underspecified for animacy. Like for other proforms already mentioned, also here, the literature mentions humanness as the relevant trigger, confirmed in the corpus of Neuburger (2008) where all indefinite proforms[human] are *a*-marked. The overall picture from the QIab survey was of a strong preference for *a*-marked exponents (Tables 45 and 46), with certain differences between lexical items. The *a*-marking of lexically specified items resulted in a higher acceptance rate than *a*-marking of underspecified ones. The case of *unu*[human] was particularly noteworthy. The preferred option here was *a*-marking.Additionally, five out of 46 informants offered a partitive construction without *a*-marking, *ne* partitive + ø + *unu*; even more informants offered an alternative partitive construction with the [animate] *unu* (Table 48), where *unu* referred to one specific dog among other dogs. Here *a*-marking was disfavoured, and the bare *unu* wasthe preferred option. According to Körner's (1987) *Korrelative Sprachtypologie* ('Correlative Language Typology'), languages with partitive systems, e.g. French or Italian, do not show DOM systems. In Corsican, the indefinite pronoun *unu* seems to trigger a partitive construction with the particle *ne.* Corsican seems to be a language that has both structures, *ne*-constructions as well as DOM. But, as the examples show, within single sentences DOM seems incompatible with *ne*. Hence, *ne* and the *a*-marker do not co-appear.

Table 47 shows the results from testing the impact of verb mood – indicative versus subjunctive – on *a*-marking. The ratings of the *a*-marked *unu*[human] did not differ notable between indicative and subjunctive verb mood.

Finally, for the negative quantifiers *nimu* and *nisunu* 'nobody', the literature refers to humanness as the relevant trigger. In Neuburger (2008) all [human] negative quantifiers are *a*-marked, but no [inanimate] one is *a*-marked. This was also confirmed by the data from the QIab survey, which showed a strong preference for *a*-marking whereby '*a*-marker + *nimu*' was rated 'good' by a very high percentage of informants (93.48%). More informants accepted the unmarked example for *nisunu*, than for *nimu* (Table 50). To test the impact of humanness, I gave also an example with *nulla*, the lexical match for inanimates: the vast majority favoured the unmarked *nulla* (Table

51).

As we have seen in previous sections, and again summarised in the last few pages, the description of Corsican DOM in the literature does not, in all respects, correspond to the data gathered for this thesis. Roughly speaking, the basic assertions in the literature were confirmed by the new data, but when considered in details, these new data provide a much more diverse picture of the distributional pattern of DOM.

7 Discussion

Previous work on Corsican (Neuburger 2008; Neuburger & Stark 2014) has centred around two main triggers for Corsican DOM, referentiality and humanness. Traditionally, accounts on DOM focus either on syntax, such as the studies presented in section 2.3, or on semantics, such as the studies presented in section 2.2. Referentiality and definiteness tend to fall under the syntactic accounts, while humanness is predominantly discussed in the context of semantic accounts. These features can, beyond their semantic and interpretative values, be related to specific structures and processes in noun phrase syntax. In Romance languages, animacy and its values do not seem to require (a) specific syntactic structure(s) of the DP (see also section 2.1).

Taking as a starting point what we know about (Romance) DOM systems in general, my interest in Corsican DOM centres on three main issues. I want to know what triggers Corsican DOM, where it appears in the DP, and why it appears there. Previous studies have revealed triggers but the descriptions were incomplete (leaving out noun classes), and the database was inconsistent and not reliable (see also Chapter 4). I enlarged the empirical database and tested the statements from the literature, in particular regarding the relevance of humanness and membership of certain noun classes. For a summary of the results, see section 6.5. Beyond humanness, I wanted to test the often-made claim (described in the literature and also observed in Neuburger 2008) that the *a*-marker is not compatible with the definite article.

The following chapter is divided into three main sections. The first of these, section 7.1, deals with the features and hierarchies presented in section 2.1; I describe Corsican DOM against these features and hierarchies. In section 7.2 I look at the *a*-marker and the definite article. I analyse Corsican DOM with reference to the accounts presented in sections 2.2 and 2.3, i.e. disambiguation and highlighting accounts for semantics, and dative and accusative accounts for syntax. I also approach the distribution of Corsican DOM from the perspective of split DP-structure, as presented in section 2.3.1. I take a closer look at the 'problematic cases' of *a*-marked indefinite and negative quantifiers (section 7.2.4) and how case assignment, i.e. *a*-marking, takes place (section 7.2.5). I conclude the chapter in section 7.3 by answering my research questions as set out in section 4.8.

7.1 Corsican DOM, features and hierarchies

Humanness is a reliable trigger for *a*-marking in Corsican. With regard to noun classes, *a*-marked proper nouns[human] and *a*-marked proforms[human] were chose as the best option by a large number of informants. The acceptance of *a*-marking is more consistent for human proper nouns[human] and

proforms[human] than for animate or inanimate nominals. The picture changes if the proper noun appears with a definite article: then the unmarked options are preferred. The results from my survey are summarised in Table 53:

Table 54: Percentage of proper nouns rated good' (based on raw data from Tables 21, 22, 24, 25, 26 and 28)

à + proper noun	*good (%)*
proper noun[human+] (religious figures) proper noun[human] proper noun[animate] proper nouns[inanimate] (months and weekdays, toponyms, brand names)	42.27 83.30 60.71 33.46
à + **definite article** + proper noun	
proper noun[human+] (ø proper noun[human]) proper noun[animate] proper noun[inanimate]	9.27 _[73] 55.56 4.65
ø + **definite article** + proper noun	
proper noun[human+] (ø proper noun[human]) proper noun[animate] proper nouns[inanimate]	38.14 _[74] 19.44 95.35

For the class of common nouns, only DOs with human referents were tested. All examples included some prenominal element – either a definite article, or a determiner (such as a demonstrative or a negative quantifier).

All the common nouns tested referred to humans. Nevertheless, *a*-marking was not the favoured option, see Table 54. Common nouns with an *a*-marked prenominal negative quantifier resulted in higher percentages of 'good' ratings than common nouns with definite articles or demonstratives. This regards the item *nisciun* 'no.one (nobody)'. The item is lexically specified as [human], i.e. it can refer to humans only. Maybe this is the explanation for the higher acceptance of *a*-marking here compared to common nouns with definite articles; negative quantifiers[human] in pronominal function, i.e. proforms, do also result in high numbers of 'good' ratings (see Tables 49 and 50, above) compared to other proforms[human], e.g. the indefinite proform *ognunu* 'everyone' (see Table 42).

---

[73] The combination '*à* + definite article + proper noun[human] was not tested in the survey.
[74] The combination 'ø + definite article + proper noun[human] was not tested in the survey.

(59) QIab.43 *Ùn*   *vegu*   *à/ø*   *nisciun'omu*

Not   see.1SG   DOM/ø no.one_man

'I do not see no man/I don't see anybody'

Table 54 provides the results for rating of options as 'good' (from Table 35) as percentages:

Table 55: Percentage of common nouns rated 'good'

	good (%)
à + definite article + common noun[human]	12.22
à + demonstrative + common noun[human]	26.83
à + negative quantifier + common noun[human]	33.33
ø + definite article + common noun[human]	86.67
ø + demonstrative + common noun[human]	68.29
ø + negative quantifier + common noun[human]	40

With regard to humanness, I also distinguished between different ways that humanness is *specified*: for some items, humanness is lexically specified, it bears, lexically or inherently, the feature value [human]; for others, humanness is underspecified, the value is given by some anchoring or linking process on the discourse level.

Informants' evaluations of indefinite proforms[human] diverged only slightly with respect to these different kinds: *a*-marked indefinite preforms non-lexically specified as human were rated #good' less often (64.44%, see Table 56) than their lexically specified counterparts (77.68%, see Table 55).

Table 56: Percentage of indefinite proforms (lexically human) rated 'good' (from data in Table 45)

qualchissia 'anyone', *qualchidunu, qualcunu* 'someone'	*good (%)*
à + indefinite lexically [human]	77.68
ø + indefinite lexically [human]	21.97

Table 57: Percentage of indefinite proforms (non-lexically human) rated 'good' (from data in Table 46)

unu 'one'	*good (%)*
à + indefinite non-lexically [human]	64.44
ø + indefinite non-lexically [human]	26.97

I also tested this distinction for universal quantifiers. Here, the difference between lexically specified as [human] and underspecified for animacy produced very different results: *a*-marked *tutti* was chosen more often as the best option (compare Tables 57 and 58).

Table 58: Percentage of informants rating universal quantifier lexically human *ognunu,a* 'good' (from data in Table 42)

ognunu,a 'everybody'	*good (%)*
à + universal quantifier lexically [human]	64.44
ø + universal quantifier lexically [human]	44.44

Table 59: Percentage of informants rating universal quantifier non-lexically human *tutti* 'good' (from data in Table 43)

tutti 'all'	*good (%)*
à + universal quantifier non-lexically [human]	93.33
ø + universal quantifier non-lexically [human]	8.00

This difference can neither be explained in terms of actual animacy values nor in structural terms; both items can function as proforms and appear in a prenominal position. Hence, lexical humanness did not automatically result in higher acceptance levels.

In addition, there was no difference of the animacy-value between items such as *Battistu* on the one hand, and *l'omu* 'the man' on the other, both items have human referents. Also, in structural terms, both are considered to be DPs, and not to be NPs, nor other potentially intermediate phrases within a split DP-structure (see section 2.3.1). Why should the one, Battistu, be *a*-marked in DO-function, while the other, *l'omu*, is not? Previous studies have argued that the definite article *prevents* or *blocks* the *a*-marking (see section 4.3). If the presence of the definite article influences the *a*-marking, then we must assume that Corsican DOM is a multifactorial system, with more than one factor influencing the marking pattern.

In section 2.1 we saw the hierarchical organisation of nominal items according to noun classes and semantic properties. The relative position of a single NE within the hierarchy allows for predictions about other NEs included in the hierarchy. Most commonly the scales are conceived as implicational scales. If the phenomenon in question, here DOM, affects a certain nominal class within the hierarchy, items of other types with a 'higher' value of the feature, will be affected as well; the further left within the scalar representation a nominal is classified, the more likely it is to get *a*-marked, and vice versa (cf. Aissen 2003: 436). In other words, if nominals[animate] are marked by DOM,

nominals[human] also get *a*-marked; if proper nouns are marked, personal pronouns also bear the marking.

Hierarchies, as presented and described in section 2.1.2, are used to describe DOM-patterns in different language varieties not only in terms of semantic values, but also, and predominantly, in terms of nominal classes.

As I have noted before, Corsican DOM does not seem categorial in absolute terms; the forced choice test neither led to clear-cut results in terms of numbers nor in terms of a distributional pattern. Hierarchical representations do not describe the Corsican data. If we compare the data from my survey to the hierarchy proposed by Laca (2006) for Spanish (see Table 4), we find 'good' rated examples within *all* classes defined in Laca's hierarchy. Laca's classes are based on semantic features and their interpretations, but do not consider the syntactic structure of nominals, i.e. the structural differences of proforms and full lexical nouns. The other hierarchies and their structures, as presented in section 2.1.2, are based on semantic features only, and do not regard the syntactic structure of the different nominal classes. Hence, DOM-systems affected or determined by syntactic structure are poorly – or not at all – represented, by these hierarchies.

To sum up, the data from the QIab survey confirm the descriptions in the literature, in that humanness of the nominal's referent has a strong impact, but to a diverging degree between different noun classes and sub-classes. For the *a*-marking of proforms, humanness is required. For the class of proper nouns, *a*-marking is preferred for human referents, but not excluded for inanimates, such as toponyms. Humanness seems to be overridden by the presence of the definite arcitle, for those noun classes that allow for it (i.e. common nouns and certain proper nouns); when nominals bear a definite article, additional DOM–*a*-marking is disfavoured.

7.2 Corsican DOM and the definite article

I approach the incompatibility of the *a*-marker with the definite article from two directions: the semantic interpretative side and the syntactic–distributional side, building on the accounts on DOM presented in sections 2.2 and 2.3.

In section 2.1 I have given working definitions of the semantic concepts of definiteness, specificity and individuation. These are interrelated, in that definiteness and specificity (and other features) contribute to individuation. Definiteness may be represented in the syntactic structure of DPs, relating definiteness to specific syntactic positions, processes and values. The treatment of specificity, however, is less clear-cut. In von Heusinger and Kaiser (2003), specificity is defined in

terms of discourse linking (for details, see section 2.1). Turning this idea around, we may postulate that discourse linking results in (one form) of (syntactic) specificity. This may be interesting for certain items; for example, proforms may be defined as [specific] due to being linked to the discourse. Their feature values are specified by the features of the nominal they refer to; for example, the animacy value of the universal quantifier *tutti* 'all' is given by the animacy value of the nouns to which it refers. See the following example (60):

(60)   *[Sgiò Baldoni, Petru, Anna è Maria]ᵢ/ₕᵤₘₐₙ anu riparatu a mo vittura. Devu ringrazià à tutti]ᵢ/ₕᵤₘₐₙ.*

        'Mr Baldoni, Petru, Anna and Maria fixed my car. I need to thank them all.'

Other proforms have invariable, lexically given feature values that are equivalent to the values of the full lexical noun they are linked to; for example the personal pronoun *ellu* 'him' bears [human] lexically or inherently, as does the noun to which it is linked.

    The relevance of specificity in terms of discourse linking is also reasonable for kinship terms and bare nouns. With regard to kinship terms, those which express a kinship relation of first or second degree and which are related to the participants, i.e. 1st or 2nd person, appear without a definite article. In DO-function they get *a*-marked. Bare common nouns, by contrast, never bear *a*-marking; they refer neither in a specific nor a definite way. They are not discourse linked but allow for a generic interpretation, as we can see in example (61).

(61)   *Omu si dumandava cum'elli evianu fattu à truvà,*

    *è   sopratuttu   à   **tene   ø   moglie [...]***

    and   above.all   to   **keep.INF   wife**

    'One asked oneself how they managed to find, and above all to keep a wife [...]' (Neuburger 2008)

Positive values of definiteness or specificity may result in overt marking. In the example above, this can trigger the use of a definite article. This applies to full lexical nouns in different syntactic functions. For nominals in DO-function without a definite article, the *a*-marker may fulfil these functions.

    We may then assume that individuated, referential noun phrases in DO-function require overt, i.e. morphological, marking for individuation and referentiality. The marking requirement can be fulfilled by a definite article, whereupon *a*-marking is neither required nor allowed. Noun phrases which do not allow for definite articles, such as proforms or (the majority of) proper nouns, get *a*-

marked to morphologically express individuation and referentiality. No double morphological marking is allowed. We may interpret this interdiction in semantic–interpretative terms, and apply Zamparelli's (2000) redundancy rule, given in structure (11):

(11) **Redundancy:** two functional words F_i, F_j within the same DP give an impossible representation if the meaning of F_i entails the meaning of F_j or vice versa. (Zamparelli 2000: 125, (333))

In terms of the redundancy rule in (11), the co-appearance of the two functional elements – in our case the definite article and the *a*-marker – would only be possible when each element adds something to the meaning of the other. In Zamparelli's example [[*the two*] *boys*], the definite article 'the' adds definiteness, and the numeral 'two' cardinality. If we interpret both the definite article and the DOM-marker *a* as a means of marking referentiality, definiteness and individuation, none of the elements add meaning to the others. Hence, in semantic terms, the incompatibility of the definite article and the *a*-marker may be explained by the avoidance of redundancy.

In syntactic terms, the definite article and the *a*-marker may compete for the same syntactic position. This may explain their complementary distribution.

To approach the syntactic incompatibility of definite article and *a*-marker, I will focus on the left edge of the DP. As we have seen in section 2.3.1, authors of different split DP-accounts consider the left edge of the DP most relevant for any processes related to the discourse, such as case marking. This includes both referentiality and individuation. The left edge comprises the DP-level(s), including head and specifier(s) (and complement), as well as other functional layers located next to it, e.g. KPs, or *n*Ps, if we assume a structure like [KP [DP [*n*P […]]]]. The features which we have discussed as relevant for *a*-marking, such as definiteness, referentiality, specificity, indexicality and the like, are associated with the DP-layer of the NP. From a syntactic point of view, the DP-level (or domain, if we include neighbouring layers) seems relevant for Corsican DOM in terms of distribution of functional elements and related features.

In the following paragraphs, I also look at animacy from a syntactic perspective.

Most syntactic accounts presented in section 2.3, elaborating DOM in Spanish, unify definiteness and animacy within one account and one structure, and assume an interplay of both features in terms of *a*-marking. The status of definiteness as a driving force for structures and processes in syntax is undeniable. Animacy is considered an ontological feature (see section 2.1) and

is traditionally perceived as a lexical, inherent feature of nominals. Cross-linguistically, it influences morphology in various ways, for example in terms of (verb or noun) inflection or lexical choice (choice of prepositions, choice of proforms). Nevertheless, it is not considered a genuine morphosyntactic feature determining processes and structures within the derivation.

In the approaches presented in section 2.3, animacy (humanness) and specificity are always 'negotiated' together in terms of syntactic structure. But *a*-marking behaves differently depending on which of these factors is driving in. With respect to humanness, the *a*-marker neither inherently bears a [human] value nor adds any animacy value to the noun. Humanness is provided by lexical information inherent to the *a*-marked nominal item, or by the context.

Regarding specificity or definiteness, *a*-marking (may) contribute to the interpretation of a nominal as specific and/or definite. For example, in Spanish the *a*-marked indefinite DO may be interpreted as specific; in Turkish too, the case marker on DOs result in a specific interpretation of the nominal (cf. De Swart 2007: 181). Also, according to Jaeggli (1982) and Diesing (1992), in Spanish the *a*-marker may trigger a specific interpretation of the nominal in question (for more detail, see section 2.1). This is not the case in Corsican: *a*-marking does not result in specific interpretations of the nominal (i.e. no difference was discernible from the choices of the informants).

Beyond definiteness and referentiality provided compositionally by different items in the syntactic structure, there are certain noun classes, such as proper nouns, which are described as inherently definite and referential. They are syntactically classified as DPs (at least since Longobardi, 1994), but for these nominals, no additional morphological marking (e.g. a definite article) is required to establish reference. Hence, definiteness and/or specificity appear both as the result of syntactic processes as well as inherent values.

Why not assume this kind of dual approach for animacy or humanness as well, and 'translate' animacy into a (morpho)syntactic feature? Section 2.3 has discussed various approaches aimed at representing animacy in the syntax of the DP. One option, for integrating animacy in DP-structure, is a (nominal) [person] feature.

Rodríguez Mondoñedo's (2007) [person] feature is not part of the lexically determined feature bundle of the nominal, but instead is triggered by humanness of the nominal's referent. It constitutes the 'morphosyntactic counterpart' to ontological humanness. The [person] feature has syntactic consequences: first it triggers an extra projection, *n*P, above NP; then it determines the specificity value of D and its projection. Giusti (2015) and Wiltschko (2014) also consider a [person] feature related to the nominal, without being part of its lexical information. What triggers [person] remains unclear, but they locate it at the left edge of the DP-structure, and relate it to the concept of

198

individuation. I come back to the idea of a [person] feature at the end of the current section (7.2).

The previous section (7.1) was dedicated to definite articles and the *a*-marker in Corsican. I reiterated the most relevant results from the database, and considered the incompatibility of the definite article from both the semantic and syntactic angles. The incompatibility of the definite article and the *a*-marker in Corsican may be explained in syntactic–distributional terms – both compete for the same position, and/or semantic–interpretative terms, both provide information about referentiality and individuation (with reference to the redundancy rule of Zamparelli 2000, see (11), above). Animacy does not seem to be relevant here; the article system in Corsican is not sensible to animacy – for example Corsican does not provide different articles with respect to animacy values, unlike languages such asCatalan. But the dual 'nature' of definiteness, i.e. aspects of semantics and syntax, has given rise to the idea that animacy might also be translated into morphosyntax, complementing its ontological nature, by assuming a [person] feature, somehow related to animacy as proposed by Rodríguez Mondoñedo (2007), Giusti (2015), and Wiltschko (2014).

To understand better the incompatibility of the definite article and the *a*-marker in Corsican, and to pursue the idea of a dual nature of humanness (animacy), I refer to the semantic accounts presented in sections 2.2.2 and 2.2.3, as well as the syntactic accounts from sections 2.3.2 and 2.3.3.

7.2.1 *Corsican DOM and semantic accounts*

The semantic accounts presented in sections 2.2.2 and 2.2.3 approach DOM (in Romance) in terms of disambiguation, highlighting or transitivity. The authors argue either that morphological marking allows for disambiguation between morphologically (and semantically) identical nominals with different syntactic functions (i.e. subject and object function); or that a high degree of transitivity (according to certain criteria and parameters, see Table 6) requires or results in morphological marking of the DO nominal. Both of these lines of reasoning have a longstanding tradition in studies on Romance DOM. So it stands to reason that Corsican DOM should be analysed from these perspectives as well.

In Corsican, potentially ambiguous context due to similar semantic values of the nominals in subject and DO-function is 'resolved' by other means than *a*-marking, for example, word order (e.g. SVO), morphological form (e.g. of personal pronouns) or world knowledge (e.g. regarding animacy). These contribute to disambiguation in a much more straightforward way than *a*-marking. Disambiguation does not seem a good candidate to explain Corsican DOM.

In terms of highlighting or markedness, I have referred to Hopper and Thompson's (1980) transitivity hypothesis and in particular to the concept of individuation: for nominals to function as

a DO, a high degree of individuation is the preferred property (see section 2.2.3, cf. Hopper & Thompson 1980). Individuation includes animacy, and also referentiality (which in turn include definiteness and specificity. DO-nominals in Corsican, that are highly individuated by [human], [definite], [specific] values get *a*-marked. Thus, Corsican DOM may be the result and/or may contribute to the transitivity of sentences. But not all of these features and values are equally relevant to all noun classes in Corsican: humanness seems to be the most relevant feature for *a*-marking of indefinite or negative proforms, neither of which refer to individuals nor are they referential. Specified as [indefinite] and [non-specific], indefinite and negative preforms are nevertheless consistently *a*-marked if they have human referents. With regard to [definite] and/or [specific] DOs, only certain *a*-marked subclasses are consistently chosen by the informants, e.g. (certain) toponyms.

Hence, transitivity may be cited as evidence for some Corsican DOM patterns, in terms of Hopper and Thompson's (1980) parameters, even though the relevance of values and features vary with the nominal classes to which the respective items belong.

The semantic accounts in terms of disambiguation and transitivity do not explain why *a*-marking occurs in Corsican. Additionally, these accounts do not provide any information about situations such as the incompatibility of the definite article with the DOM marker. So, in the following section I look at Corsican DOM from the perspective of the syntactic accounts on DOM presented in section 2.3.

7.2.2   *Corsican DOM and syntactic accounts*

In section 2.3 I presented syntactic accounts on DOM. First, I gave a detailed insight into split DP-structures and feature distribution within DPs. Then I presented some of the best-known syntactic accounts on Romance DOM. I will start my syntactic analysis with the so-called dative and accusative accounts (sections 2.3.2 and 2.3.3).

In Corsican, datives are all marked by *à*. This includes lexically–semantically driven datives as well as structurally driven cases, such as IOs in ditransitive constructions. I assume that dative can be both a structurally and lexically driven case (see also section 2.3.2, following work of Baker 2015 and Næss 2012). With regard to accusative case, I adopt the common assumption that accusative case is assigned structurally and, in a broad sense, emanates from the verb.

If I want to explain the incompatibility of the definite article and the *a*-marker in terms of 'competition' for the same syntactic position, dative accounts on DOM, as described in section 2.3.2, are interesting. The authors of these accounts deduce, from ditransitive constructions where dative is consistently *a*-marked but DOM-*a*-marking is often omitted, a syntactic structure with a single

position for *a*-marking. Here, the *a*-marker is obligatorily assigned to dative objects; accusative objects, meanwhile, remain unmarked, regardless of any semantic or syntactic DOM-triggers. Beyond ditransitives, another source of dative accounts are Spanish existential constructions with the verb *haber* (for details see Rodríguez Mondoñedo (2007) in section 2.3.2).

For Corsican DOM, I did not test for *a*-marking of DOs in ditransitive or existential constructions. But idiosyncrasies of the dative accounts presented in 2.3.2, i.e. a single syntactic position for both dative and accusative object and/or a single case assigning process for both objects, do not seem reasonable from another point of view: in none of the dative accounts does case assignment seem sensitive to the syntactic structure of the DP, in particular the presence of a definite article. In Corsican, this sensitivity to the structure of the nominal distinguishes the *a*-marking of accusative objects and dative objects. Dative *a*-marking appears on any nominal regardless of its syntactic structure, including prenominal elements such as definite articles. For accusative objects the same does not hold true; the *a*-marker does not appear when the DO bears a definite article.

The dative accounts presented do not provide options for integrating the described sensitivity of Corsican *a*-marking. But they do allow for insights into the structure of DPs. Torrego (1998) argues for a dative interpretation for *certain a*-marked DPs in Spanish, but not all of them; not all *a*-marked DPs in (direct and indirect) object function behave the same. She specifies three structurally different *a*-marked objects (see Table 9, section 2.3.2). Two constructions are classified as PPs, where the *a*-marker is hosted in P°. The third *a*-marked construction is a DP. Torrego remains very vague about this *a*-marked DP, with regard to both its location in the derivation and its internal structure. It remains unclear where the *a*-marker is hosted, e.g. in SpecDP or D° (see also Figures 4 and 5, section 2.3.2), whether the *a*-marked construction has to move out of VP, and if so, where it gets the case marking. Nevertheless, Torrego acknowledges structural differences among *a*-marked objects by distinguishing between *a*-marked DPs and PPs.

For Rodríguez Mondoñedo (2007), *all a*-marked DOs must be raise to a dative phrase located above *v*P, with no structural differences (see Figure 6, section 2.3.2). Regarding the internal structure of the *a*-marked object, Rodríguez Mondoñedo (2007) classifies it as a KP with a complex internal structure [$_{KP}$K [$_{DP}$D [$_{n}$Pn [NP]]]]. K is realised as /*a*/ if certain feature requirements are met, i.e. the DP bears [human] and [person], in NP and *n*P respectively (see Figure 7, section 2.3.2). As mentioned above, Rodríguez Mondoñedo (2007) assumes that [human] NPs trigger a [person] feature in their nP layer, which in turn is relevant for the specificity of the D-head and, finally, for *a*-marking. Hence, the [human] and [person] features do not change the categorial status of the KP/DP, but they have a morphosyntactic impact. How this KP is related to the PPs and the dative/accusative

assignment process remains unclear.

The third dative account presented is from Ormazabal and Romero (2013). Like Rodríguez Mondoñedo (2007), they assume one single case assignment process for dative and accusative objects. In ditransitive constructions, only one of the two objects is case checked and marked, while the other remains altogether caseless, the DO. As proper caselessness would cause the derivation to crash, Ormazabal and Romero (2013) propose a repair-mechanism specifically for accusative objects. How the process of accusative *a*-marking works in transitive contexts remains vague and nontransferable. Probably the caseless (i.e. unmarked) DOs are part of some incorporation structures, as proposed by López (2012).

None of the (dative) structures offered by Torrego (1998), Rodríguez Mondoñedo (2007) or Ormazabal and Romero (2013) account for the incompatibility of the definite article and the *a*-marker. This comes as no surprise, as dative case assignment (in Spanish) is not sensitive to the syntactic structure of the nominal.

Beyond the issue of case assignment, Rodríguez Mondoñedo (2007) introduces a split DP, a complex KP-structure, [$_{KP}$K [$_{DP}$D [$_n$n [NP]]]]. The functional layer KP above DP relates Rodríguez Mondoñedo's (2007) account structurally to the 'accusative accounts' presented in section 2.3.3. In these accounts, the functional layer hosts, one way or another, the *a*-marker in its head position. For Brugè and Brugger (1994) it is a Function Phrase, FP. López (2012) assumes a Case Phrase, KP, where the head, K°, is realised as /*a*/ by late vocabulary insertion.

Taking again the complementary distribution of the definite article and the *a*-marker as a starting point, Brugè and Brugger's (1994) account does not provide an explanation, in simple terms, of competition between these two items for the same syntactic position. They extend the DP by an FP (see Figure 17 in section 2.3.3), providing, in F°, a position for the *a*-marker. The phrase below, a DP, offers a position for the definite article in D°. By offering two positions for the *a*-marker and the definite article, both items may be hosted in the syntactic tree, if no further conditions apply.

López's (2012) approach (on Spanish DOM) also offers no solution to account for this Corsican peculiarity. In terms of López's syntactic structure, full lexical nouns and proper nouns[human] realised with a definite article project into a DP, then into a KP, and provide all feature values required for K to be realised as /*a*/. But in Corsican, no *a*-marking occurs here. The system presented in López (2012) does not provide a tool to explain how (and why) the overt realisation of the definite article prevents the overt realisation of K°, i.e. *a*-marking.

Neither of the accusative accounts reviewed provide tools to account for the Corsican

peculiarity. Nevertheless, certain aspects of their approaches are interesting for a better understanding of syntactic structures and the processes within: namely, the idea of *a*-marked DOs being KPs or FPs, but not PPs; and the postulation of a functional layer above DP to host *a*-marking, i.e. FP or KP, and a single case checking/assigning process for all accusative objects, DOs. Neither account tackles problems such as the ability of the definite article to block/prevent *a*-marking.

7.2.3   *Corsican DOM in a split DP*

For most of the syntactic accounts presented, an additional functional layer beyond the DP is required for DO nominals to get *a*-marked, be it the functional layer FP or KP (cf. Brugè & Brugger 1994; López 2012) or the 'dative phrase' of Rodríguez Mondoñedo (2007). Such an approach follows naturally from Abney' (1987) introduction of the DP-structure. The DP-layer is considered a prerequisite for argumenthood. In additional to argumenthood, the projection of a D-layer also shifts nominals from predicates to individuals, i.e. from $<e,t>$ to $<e>$ type (cf. Alexiadou et al. 2007: 64).

With regard to argumenthood, Déchaine and Wiltschko (2002) demonstrate that, (at least) regarding proforms, a DP-projection is not required for argumenthood. They distinguish three categories of proforms according to semantic and syntactic criteria, i.e. pro-DPs, pro-φPs and pro-NPs (for more detail, see section 2.3.1). They show that pro-DPs always have argument status, but, vice versa, not all arguments are necessarily DPs. Also, pro-φPs can be arguments. Hence, the projection of a DP-layer might not be required with regard to argumenthood, and can probably be suspended in contexts where the referential properties of the nominal do not trigger such a structure.

Concerning the type-shifting function of the DP, Fábregas (2013: 52) argues that KP and DP, in their overt realisation, share the semantic and syntactic capacity to function as both type-shifters and case markers. Zamparelli's (2000) determiner system (see section 2.3.1) provides a double DP-structure, but type-shifting occurs only at the topmost DP-layer, SDP. Only SDP noun phrases are of type $<e>$. However the projections or layers are labelled, the highest functional layer –DP$_1$, KP or SDP – is responsible for type-shifting and probably case marking, but not for argumenthood. The argument given by Fábregas (2013), that both head elements in KP and DP may function as case markers, is in line with Giusti's (2015) main assertion that the definite article is a case morpheme.

Giusti (2015) presents two arguments to support her idea: the distribution of definite articles and case morphemes are both dependent on syntactic mechanisms; and both are functional heads. As presented in section 2.3.1, Giusti claims that the definite article is a discontinuous part of N, namely, Case. In her account, the definite article does not establish referentiality automatically by virtue of its lexical information; an independent referential index, ι, is conceptualised, which is hosted in SpecDP. The referential index ι may co-occur with a definite article, but does not necessarily do so;

for example, in the case of generic interpretation of the DP, there is no ι. There are other items that bear such an index inherently, e.g. personal pronouns, demonstratives and proper nouns; these can be merged lower in the structure, but move then to SpecDP (see Figure 2, section 2.3.1).

Additionally, Giusti claims that not only is ι separated from D°, but that [person] is also separated from N (arguing that no nominal is inflected for person of its own referent). According to her, [person] is also hosted in SpecDP, the left edge of the DP. Hence, both ι and [person] are part of the left edge of the nominal domain, where they may merge with the nominal (cf. Giusti 2015: 50).

While D° and the D-element are not responsible for the referential interpretation of the nominal, i.e. neither bear or provide ι, D° hosts a unvalued feature for Case: [*u*Pred(Case)]. According to Giusti (2015: 110), [*u*Pred(Case)] allows for the nominal to be part of the next phase, to be accessible to the probe of *v*P. It is Case that allows the nominal to re-enter the next phase as part of the denotation (cf. Giusti 2015: 125). Hence, [*u*Pred(Case)] can be interpreted as a trigger to project a DP-layer. [*u*Pred(Case)] and a D° which are not automatically related to referentiality, facilitate the projection of a D-layer; the nominal does not necessarily require the referential qualities most commonly tied to the projection of a DP. Items such as indefinite proforms may project into DP for case requirements.

The DP accounts of Zamparelli (2000) and Wiltschko (2014) are interesting from different points of view: Zamparelli's double DP-structure may also allow for a DP-structure for non-referential DPs, i.e. his 'lower' PDP. This DP may host indefinite proforms. Argumenthood may also be provided/tied at this level of DP-structure. Similarly, Déchaine and Wiltschko (2002) show that the highest projection of the nominal domain, i.e. the D-level, is not required for argumenthood; proforms (at least) may project into a φP, the projection below the D-level, and function as arguments. Hence, both accounts separate argumenthood from the highest projection in the DP. This yields two advantages: they allow for items to be arguments irrespective of DP status; and consequently, they allow for items to be in these positions, that are difficult to account for under different DP approaches, e.g. indefinite or negative quantifiers (for more on indefinite and negative proforms, see the following section (7.2.4)).

To account for the incompatibility of the definite article and the *a*-marker in Corsican, Giusti's (2015) idea that the definite article is proper case-morphology is central. This idea, applied to Corsican data, allows for the 'equalisation' of the definite article and the *a*-marker under the single label of 'case marker'. Either the definite article or the *a*-marker can fulfil the case requirement for nominals. In addition, an independent referential index, ι, separated from D-elements (i.e. the definite article) and the D° positions, allows for projection of a determiner phrase without the need for

referentiality: this allows for *a*-marked indefinite proforms[human] in a DP-structure. Furthermore, if [person] is also an independent feature, rather than part of the feature bundle of the nominal, its presence may be triggered by the humanness of the referent of the nominal in question. Hence, humanness may be translated into a morphosyntax feature and allows for morphosyntactic processes to be sensitive to humanness, such as Corsican DOM.

7.2.4 *Corsican DOM and indefinite and negative proforms*

The *a*-marking of indefinite and negative human proforms causes problems, in both of the accounts presented in section 2.3. Rodríguez Mondoñedo (2007) and López (2012) analyse the syntactic structure of lexical items from these noun classes, namely Spanish *nadie, ninguno,a* 'nobody', *alguien, alguno,a* 'someone', and *quien* 'whom'. As we have seen, both accounts consider a functional phrase above DP, that is a KP, to host the *a*-marker. For indefinite, non-specific lexical items, such as indefinite and/or negative proforms, the projection of a D-layer as it is commonly understood is problematic. We usually assume that referentiality is related to the DP level, and definiteness, specificity, and/or individuation of the nominal contribute to the nominal's referentiality. The indefinite and negative proforms are associated with [indefinite] [non-specific] values, and as a consequence they should not be able to project into a DP. How, then, can these items have a KP (to get *a*-marked), but apparently no DP? Or, if they do have a DP, how is it triggered?

Rodríguez Mondoñedo (2007) analyses the negative quantifier *nadie* 'nobody', the indefinite proform *alguien* 'somebody', and the interrogative *quien* 'whom'. All items inherently or lexically bear the feature [human]. Morphosyntactically, Rodríguez Mondoñedo decomposes the lexical items into their single morphemes and assumes two different D-heads to account for the distribution of DOM in Spanish: one of the D-heads is underspecified for specificity, hence [± specific], while the other is marked as [non-specific] (for details of Rodríguez Mondoñedo's 2007 account, see section 2.3.2). For Rodríguez Mondoñedo, indefinite and negative proforms project into a *DP (DP with an asterisk), with a *D-head which is [non-specific] by definition. The nominals hosted in such a *DP do not provide and also do not require specificity for the DP level.

The internal morphological structure of these items is represented in the following Figure (14):
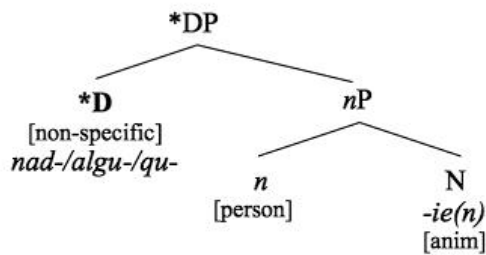
*Figure 14: Morphological structure of *DPs with* nadie, alguien *and* quien *(Rodríguez Mondoñedo 2002:202)*

By hosting lexical material in the functional head D, Rodríguez Mondoñedo (2007) contradicts established approaches. He does not explain this issue any further.

On the clausal level, the internal organisation of the lexical items is not visible, but the *DP appears as a feature bundle, such as that given in Figure (14). The features as such are accessible for syntactic processes: the [person] feature triggers the realisation of */a/* in K.
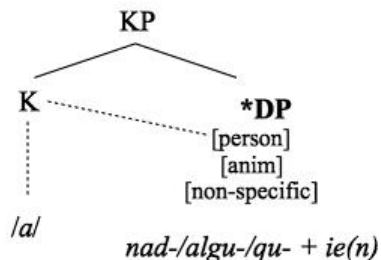
*Figure 15: *DP nadie/alguien/quien within the KP (cf. Rodríguez Mondoñedo 2007)*

Rodríguez Mondoñedo (2007) assumes another D-head underspecified for specificity. This D can be either valued as [specific] or as [non-specific]. Rodríguez Mondoñedo's (2007) argumentation for distinct D-heads with regard to specificity is hard to follow.

With regard to the general relation between specificity and DP, we know that not all DPs are interpreted as [specific]; evidence comes from nominals that bear a definite article and are interpreted generically. An underspecified D-head allows for different interpretations *and* it conforms to our

expectations about indefinites and negative quantifiers by being non-specific in semantic terms.

Nevertheless, the DP-layer for these elements is controversial from a semantic *and* a syntactic point of view, for the following reason: if specificity (and definiteness) are not provided by indefinite and negative proforms[human], how can these items project a DP? In other words, which features trigger the DP? Like Rodríguez Mondoñedo (2007), López (2012) works with two distinct D-heads to account for the *a*-marking of indefinite proforms[human]. While Rodríguez Mondoñedo argues on the basis of specificity, López draws on humanness. Humanness seems more promising, as it appears to be less counterintuitive for the human indefinite and negative proforms. López assumes one D-head marked as animate, i.e. D°[anim], and a second one which is underspecified for animacy, i.e. D°[±anim]. The two different heads may host distinct items. Some of these items are inherently and/or lexically animate, such as *nadie* or *alguien*; these are hosted in a DP[anim]. Others, which are underspecified for animacy, such as *ningún-* and *algún-*, are hosted in a DP[±anim]. They get their animacy value from the noun to which they refer (see López's (2012) syntactic structure in Figure 16):
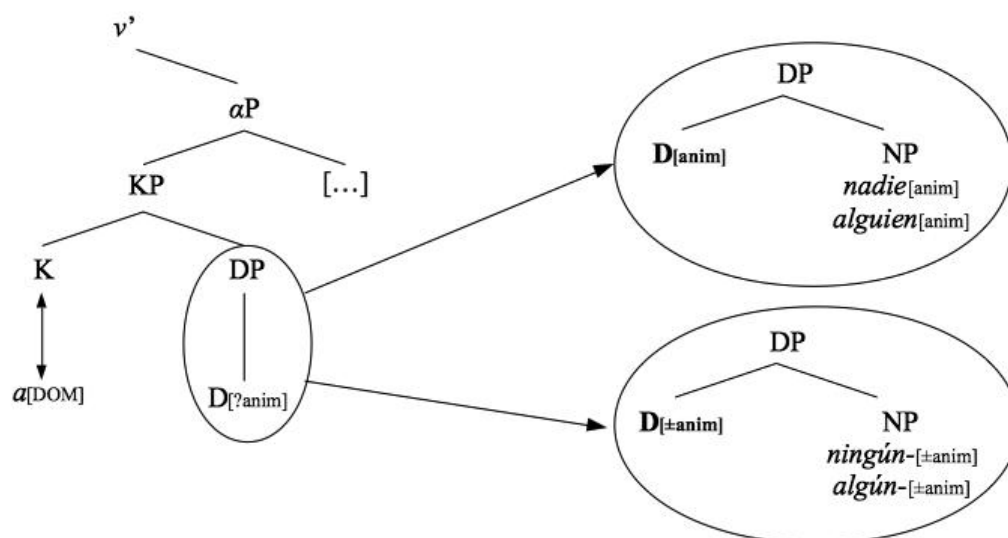
Figure 16: López's (2012) DPs with inherent versus underspecified animacy values

No conceivable advantages arise from the assumption of a double set of D-heads, i.e. from assuming, in addition to the underspecified D-heads, either a D-head marked as [non-specific] (Rodríguez Mondoñedo (2007) or a D-head marked as [anim] (López 2012). These underspecified D-heads get their values from the nouns hosted in their NP. No difference between inherent and 'ascribed' values

are observed, at least for Corsican DOM (e.g. no difference in acceptance, between proforms lexically marked as [human] versus those structurally marked as [human]). But a DP which is sensitive to animacy is interesting from different points of view; it allows for the projection of DP, regardless of features traditionally associated with the projection of this layer, such as referentiality or definiteness. In turn, the projection of the DP per se confirm traditional assumptions that a DP is necessary to project into a KP, the functional layer for case. Such an account allows for the structural representation of the interrelation between case and animacy, in particular *a*-marking and humanness, irrespective of referentiality in terms of definiteness. We may then represent the *a*-marking of indefinite and negative proforms[human] within a traditional [KP [DP [NP]]] structure.

Also, from an economic point of view, a single D-head underspecified for the above-mentioned features is the preferred solution.

Giusti (2015) offers another approach, to fulfil the structural requirements for *a*-marking of indefinite and negative proforms[human]. These requirements relate to the structural status of the preforms in question, and where to host the case marker. Giusti's (2015) [*u*Pred(Case)] feature triggers the projection of a DP, regardless of any definiteness, specificity or referentiality values. The checking relation for this unvalued feature allows for case assignment. Case can be overtly expressed by the *a*-marker. Hence, case requirements trigger the projection of a DP; in the respective noun phrase, we find nominals of different kinds, including indefinite and negative proforms. Proforms do not allow for definite articles: in case the proforms refer to humans, the case marking will be implemented in the form of the *a*-marker. In terms of Giusti's (2015) tree-structure it will appear as shown in Figure 17:
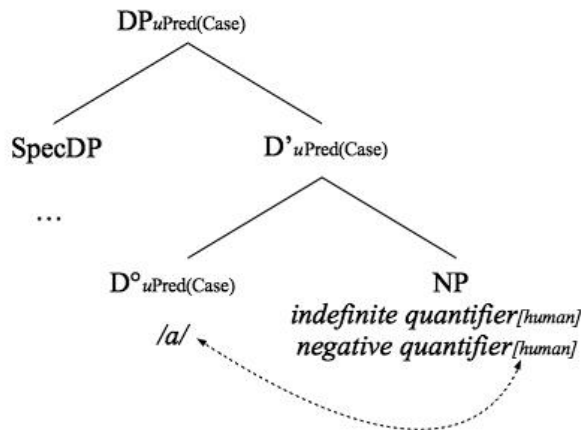
Figure 17: Giusti's (2015) DP$_{[uPred(Case)]}$

To summarise: the implementation and results of Rodríguez Mondoñedo's (2007) account do not result in coherent advantages for indefinite and negative proforms[human]. The advantages of López's (2012) account are obvious; a D-head sensitive to animacy or humanness (and independent from referentiality) may allow for indefinite nominals[human] to project into a DP. In this way they fulfil (López's, and others') structural requirements for the projection of KP, to get *a*-marked.

Giusti's (2015) DP also allows for indefinite and negative proforms. This is possible because she separates referentiality, i.e. a referential index, ι, from D°. The index is hosted in SpecDP. While for D°, she assumes the unvalued case feature, [*u*Pred(Case)]. In addition, [human] and certain indefinite and negative proforms also need to fulfil case requirements, independently of definiteness values. Hence, they project into a DP as other DO nominals do. In Giusti's terms, the definite article is overt case morphology hosted in D, and valuing the [*u*Pred(Case)] feature. As proposed before, we may extend this idea and assume that D can host case marking items in general; in Corsican this means the definite article and the *a*-marker. This allows the structure to represent the *a*-marking of indefinite and negative proforms[human].

The accounts of Rodríguez Mondoñedo and López may offer a solution for indefinite and negative proforms[human] to be hosted in a KP–DP-structure, but looking beyond these items, these accounts do not offer a solution for the Corsican incompatibility of the *a*-marker and the definite article. With this idiosyncrasy of Corsican DOM in mind, Giusti's DP offers a solution in terms of distribution – by providing, in D°, a single locus to host case marking devices, and by considering that the definite article case morphology, of either the definite article or the *a*-marker, overtly mark

case and compete for the same position. This is discussed further in the following section (7.2.5).

7.2.5 *Corsican DOM and case assignment: the marking process*

In addition to the questions of how to deal with complementary distribution, and where to host the accusative *a*-marker, we need to be specific about case assignment – about how case is assigned, and what is marked and how. Essentially, accusative case assignment draws on feature checking/valuation. López's (2012) account can derive *parts of* the distributional pattern of Corsican DOM quite well. DOs may check case in two ways. One possibility is the projection of a so-called KP, in which case they rise (from the CompVP position) into a functional projection αP between VP and *v*P. Here the DO-DP gets case checked by *v*P and, if it bears certain feature values such as [human], K° will be realised (phonetically) as /*a*/. If the DP does not bear the features, it is still case checked, but K° will not be overtly realised (for details and Figures, see section 2.3.3). (The second possibility of case checking includes an incorporation–movement process, in detail see section 2.3.3). A unified KP-structure, for *a*-marked as well as unmarked DOs, is attractive from different points of view: it recognises the *a*-marker as being the overt realisation of case; at the same time, case assignment does not hinge on the overt occurrence. But Lopez`s account does not solve the problem of incompatibility between definite article and *a*-marker.

For Corsican human nominals in DO-function I take overt case marking for granted. Following
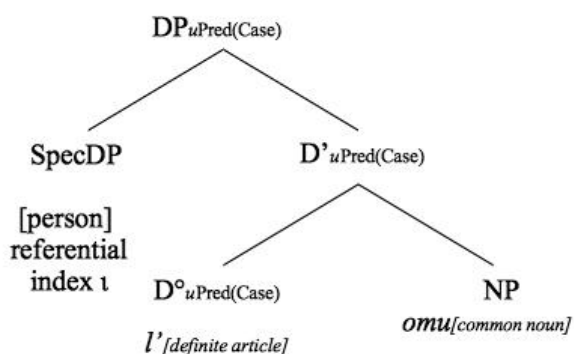
Figure 18: Case marking by the definite article: l'omu

Giusti (2015), I assume for DOs that Case can be expressed either in the form of a definite article or by *a*-marking. The structure of the nominal in question determines which element fulfils the case marking requirement. Some nouns allow for definite articles, others do not. In the following Figure (18), we see case marking by the definite article:

Common nouns, such as *omu* 'men', are assigned morphological case by the realisation of a definite article in D°. In this way, the unvalued feature [uPred(Case)] becomes valued. In addition, the animacy value [human] of the noun may trigger a [person] feature in SpecDP. And, depending on context, we may find a referential index, ι, in SpecDP. Both features contribute to referentiality.

For nominals which do not allow for definite articles, such as proforms or proper nouns, overt case marking is expressed by *à*. Again, humanness of the noun may trigger a [person] feature in SpecDP, and, depending on the semantics of the noun, also a referential index, ι. This may result in the following Figure (19):
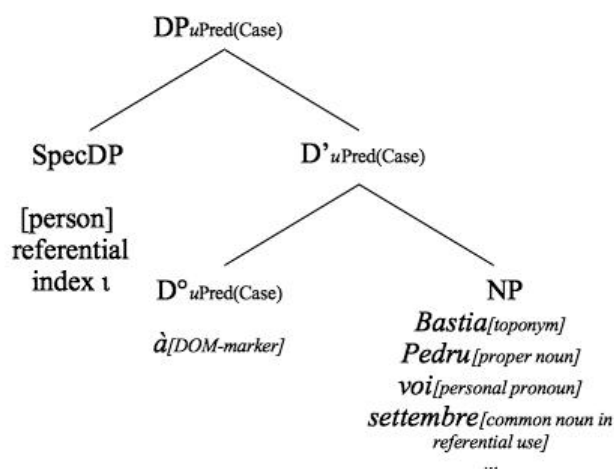
Figure 19: Case marking by the a-marker

While the unvalued case feature [uPred(Case)] triggers the projection of D, both humanness and/or definiteness seem to trigger *overt case marking* on the DO nominal – bin this way, Corsican DOM manifests itself in quite traditional DOM-terms. What remains remarkable is the double marking strategy provided by the *a*-marker and the definite article.

In summary, with the case marking systems presented above in Figures 18 and 19, we may account for different characteristics of Corsican DOM: for the *a*-marking of indefinite and negative human proforms (within a DP-structure); the *a*-marking of other nominals not allowing for definite articles (e.g. proper nouns); and the complementary distribution of the definite article and *a*-marking with full lexical nouns.

Sentences with *A Madonna, U Sciroccu* and *A Marana* (see Tables 21, 25 and 34) do not conform to this structure. These sentences are similar in that these proper nouns bear the definite article as part of the nominal. They have different kind of referents in terms of animacy: [human+] for the religious figure *A Madonna;* [animate] for the horse *U Sciroccu;* [inanimate] for the toponym

A Marana. Remarkably, these sentences resulted in quite different reactions from the QIab survey informants. While, for *A Madonna* and *A Marana*, the majority chose the option with the definite article over the *a*-marking, for *U Sciroccu* the simultaneous appearance of *a*-marking and the definite article was the preferred option. At the same time *U Sciroccu* provoked the most *alternative* responses from informants, replacing the definite article with the *a*-marker (for convenience I re-iterate the data from Tables 21, 25, and 34, in Table 59).

Table 60: A Madonna, U Sciroccu, and A Marana

Proper nouns with definite articles		total	*'good'*	*'bad'*	*?*
Iab.13a	*Prighemu à a Madonna*	46	9	8	2
Iab.13b	*Prighemu ø a$_{[DET]}$ Madonna*	46	37	2	2
Iab.13c/ alternative	*Prighemu à (*a$_{[DET]}$) Madonna*		1		
Iab.34a	*Petru s'hè coltu à U Sciroccu*	46	20	4	1
Iab.34b	*Petru s'hè coltu ø U Sciroccu*	46	7	11	0
Iab.34/ alternative	*Petru s'hè coltu à (*U) Sciroccu*		9		
Iab.33a	*...copre à La$_{[DET]}$ Marana*	46	4	6	2
Iab.33a	*...copre A$_{[DET]}$ Marana*	46	41	2	1

Informants' responses showed a clear distinction between the feminine definite article and the masculine definite article; *a*-marking of the former was disfavoured, *a*-marking of the latter as to some extend accepted. Both the feminine article and the DOM-marker were pronounced /a/, with no phonetic differences. The preference for a_{DET} *Madonna* and a_{DET} *Marana* may be explainable in terms of haplology: phonetically a_{DET} and a_{DOM} merge into one single /a/, hosted in D°. How this /a/ would be labelled, either as a definite article or as a DOM-marker, I have not investigated any further. Probably it is simply $a_{[case]}$.

With respect to *U Sciroccu,* it is probable that most of informants did not perceive or consider the definite article *u* not as an independent item, still less as a case marker. Hence, the preference for *a*-marking of *U Sciroccu* may be the result of the required case marking. If this is so, the syntactic structure may look like the tree in Figure 20:
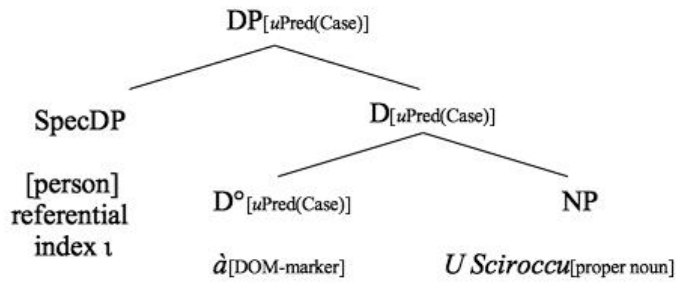
Figure 21: Case marking of U Sciroccu: à U Sciroccu

Those informants who perceived *u* as definite article either chose the unmarked option, i.e. *U Sciroccu* without *a*-marking, or offered an alternative, substituting the definite article *u* with the *a*-marker, *à Sciroccu.* Either way, *Sciroccu* is case marked. Then the following structure (Figure 21) applies, similar to that shown in Figure 20:
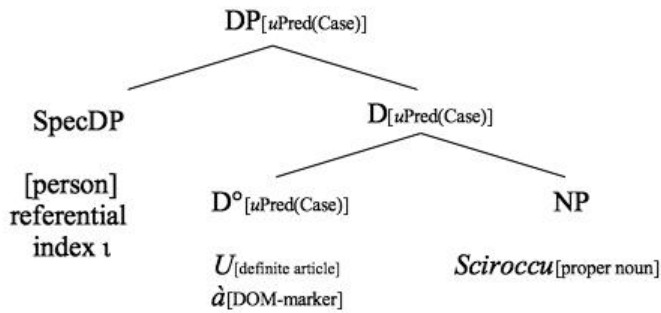
Figure 20: Case marking of Sciroccu: U Sciroccu *or* à Sciroccu

In summary, DOM in Corsican is the result of an *overt case marking* requirement for DO-nominals with human and/or referential referents. Case may be expressed either by a definite article or by the *a*-marker. The precise functional item that will case mark a given nominal depends on the nominal: nominals which do not allow for definite articles (in their usual form) are *a*-marked; nominals that do allow for the definite article are marked by said article; no *a*-marking takes place.

It remains subject to scrutiny whether all DOs in Corsican require overt case marking; and, if what case marking looks like for non-referential, inanimate nominals that are not relevant for DOM (e.g. common nouns with an indefinite article), or indefinite inanimate nominals (e.g. the negative quantifier *nulla,* 'nothing'). The latter may probably merge into some kind of incorporation structure which moves on for case requirements, similar to López's (2012) idea illustrated in Figure 12. These issues I leave for further research.

7.3 Summary of the discussion: the three research questions

I conclude the discussion by answering my research questions, as set out in Chapter 4 section 4.8 (repeated here for convenience):

1. How is humanness related to noun classes and syntactic structure, considering the *a*-marking of many DOs[human], but not all?
2. Why do the definite article and the *a*-marker not co-appear? (Why) does the definite article prevent or block *a*-marking? How can is this distribution be explained in semantic–interpretative and syntactic–distributional terms?
3. And finally, the DOM-patterns of certain lexical items such as indefinite and negative quantifiers fall outside the rules developed for other lexical patterns, both with regard to semantic and/or syntactic argumentation. By what means may they be accounted for?

Starting from the basic assumption that referential and/or human DOs need overt case marking in Corsican, I will answer the research questions as follows:

1. We have seen that humanness is the most consistent trigger for Corsican DOM, throughout all noun classes except common nouns. In the class of proforms[human] and proper nouns[human], the *a*-marked option was more often chosen by the informants than the unmarked option. The reverse of this pattern applies only partially: *a*-marked proforms[inanimate] were almost never chosen as the best option, while certain *a*-marked proper noun[inanimate], such as toponyms or month names, were consistently chosen as the best option. Here, the second trigger may apply; in terms of

definiteness, DO nominals with a high degree of referentiality need overt case marking – either by a definite article, as in the case of common nouns (Question 2, and its answer below), or, for items such as proper nouns, which do not allow for definite articles, by the *a*-marker.

2. The noun classes included in the survey are distinct in structural terms, with respect to their ability to appear with or without a definite article. Proforms do not allow for definite articles, and proper nouns only allow for them under specific circumstances – for example if the determiner is part of the proper noun itself (e.g. as a result of lexicalisation). In contrast, common nouns appear with definite articles if they are interpreted as definite, referential, specific, individualised, or indexical, and do not allow for *a*-marking even if they refer to humans. Giusti (2015) argues that the definite article is discontinuous case morphology of the noun. Given the complementary distribution of the definite article and the *a*-marker in Corsican with respect to referential and/or human DOs, as discussed in sections 7.2.3 and 7.2.5, I ascribe the case marking function to both the definite article and the DOM *a*-marker. They appear in complementary distribution because different nominals in DO-function allow for different marking strategies: some allow for definite articles, such as common nouns; others do not, such as proforms and (most) proper nouns. The definite article does not *block* the *a*-marking in a proper sense; the distribution of either the definite article or the *a*-marker is driven by the general ability of the noun to appear with or without definite articles.

   In syntactic terms, the definite article and the *a*-marker both merge into the D° position, in structures such as that given in Figure 19. D° is either expressed via the definite article or the *a*-marker, but the two items cannot not co-appear, as D° hosts only one case marker at a time. There are very few exceptions where the *a*-marker and a definite article appear simultaneously, such as *à U Sciroccu*. In this case the definite article may be lexicalised and builds part of the proper noun (see Figure 20, above).

   In semantic–interpretative terms, we can apply Zamparelli's Redundancy Rule: no two functional items may co-appear if the meaning of one comprises the meaning(s) of the other (for the original wording, see 11).

3. Neither semantic nor syntactic accounts, such as those presented in sections 2.2 and 2.3, have provided a straightforward analysis to account for *a*-marking of indefinite or negative proforms[human]. In terms of referentiality or definiteness, these items fall through the gaps because they are not referential so marking triggered by definiteness does not apply to them. In terms of current syntactic approaches to Romance DOM, these items do not easily meet structural requirements for *a*-marking. The syntactic accounts of section 2.3 require a DP-structure to build

up an additional functional phrase related to case marking; for example, KP. Negative and indefinite proforms[human] usually do not project into DPs, due to the absence of referentiality and/or definiteness. As a consequence, they have no KP-structure either. López (2012) solves this problem by offering a DP-structure that is sensitive to animacy, in which the animacy value of the negative and indefinite proform[human] triggers the projection of a specific DP and thus also the ability to project a KP, to allow for overt case marking by *a*.

In section 7.2.4, I elaborated a different structure for indefinite and negative proforms[human], where some of López's (2012) and Giusti's (2015) ideas intertwine. I assume a phrase structure in which overt case marking of accusatives can be triggered by humanness and/or referentiality. The values of these features are provided by the nominals in question, either inherently, (e.g. proper names, which are considered inherently referential) or structurally (e.g. proforms, which get their animacy value from the nominal they refer to), or may be considered definite in terms of discourse linking or anchoring (cf. von Heusinger & Kaiser 2003, section 2.1.2). Case marking in Corsican can be expressed either as a definite article or the *a*-marker, and the two items do not co-appear within a single DP. Hence, a single case marking position, i.e. D°, which is triggered by case requirements (i.e. the unvalued feature [uPred(Case)]), would be sufficient.

8 Conclusion & Outlook

This dissertation aims to investigate Corsican DOM by means of a new empirical database and an analysis in terms of current theories of Romance DOM. Through this work, I want to contribute to the field of Corsican linguistics in particular, but also to the research on Romance DOM in general. In addition to Corsican DOM, I have also provided information about the history of Corsican and its speakers and given some insight into the history of the science relating to Corsican as an independent field of linguistics.

The Corsican peculiarity, i.e. the incompatibility of the definite article and the *a*-marker, serves as a starting point to investigate the Corsican DP in detail. This detailed analysis results in new syntactic structures, which may also account for well-known problematic cases beyond Corsican, such as DOM-marked indefinite and negative quantifiers[human] in Spanish. To underpin my analysis of the data from my survey, I first introduced, in Chapter 2, some of the most conventional semantic and syntactic accounts on Romance DOM, all of which are founded on the basis of features (section 2.1). For (Romance) DOM, the most important features are considered to be definiteness and specificity, including concepts such as individuation and referentiality, as well as animacy, and humanness in particular. The features may be conceived differently. Some have semantic–interpretative value only, such as the ontological feature animacy; the ontological animacy value of a given nominal cannot be changed. Other features have a semantic–interpretative impact, but also influence syntactic processes such as definiteness or referentiality; they seem to trigger the projection of a DP.

The most longstanding tradition, among the accounts on DOM that I have presented, is found in the so-called 'disambiguation accounts' 2.2.2. Essentially, these accounts proceed from the need to disambiguate the nominals in subject and object function by morphological means, if these nominals are too similar in semantic terms (in terms of animacy and/or definiteness, and in their morphological appearance) and cannot be distinguished by other means such as word order. The basis of the disambiguation hypothesis is the assumption that subject nominals and object nominals bear certain feature values in their most common or prototypical occurrence (for discussion of prototypicality and markedness, see section 2.2.1). Subjects are assumed to be [human] and [definite], objects are [inanimate] and [indefinite]. If objects are [human] and [definite], they deviate from the basic pattern. In terms of markedness, these deviant nominals are marked. Thus, the overt morphological marking of [human], [definite] DO-nominals corresponds to this markedness based on deviation.

I elicited the data for my corpus by means of a written questionnaire, and did not compile a corpus of naturally produced language. As such, I am not able to use this data to talk about the

proportion between [definite], [human] DOs and [indefinite], [inanimate] DOs in naturally produced language and the potential need to disambiguate between subject and object nominals. Regarding the morphological appearance of DOs: as with most Romance languages, Corsican too provides specific lexical items for accusative objects such as personal pronouns and interrogatives. It also tends to follow the basic word order SVO. Both of these mechanism allow for a more effective discrimination between the two nominals than the *a*-marker would. In particular, if we take into account that the *a*-marker may even cause ambiguities (Corsican datives and locatives are also marked by *a*, and the *a*-marker is omitted in ditransitive context where the dative object bears *a*-marking), then disambiguation is not a good candidate to explain the motivation for Corsican DOM.

DOM-marking, explained in terms of transitivity (2.2.3), offers a better understanding of the markedness relation from the opposite direction: in a transitive construction, in accordance with the specific parameters pertaining to subject, object, and verb, high degrees of definiteness and animacy on the DO-noun speak of a high degree of transitivity. Hence, [human] and [definite] DOs contribute most to the transitivity of a sentence. The *a*-marking is not the result of a deviant pattern, but because the pattern fits most in terms of transitivity, it is thus a highlighting device.

Again, as my corpus does not contain naturally produced language, I cannot speak about the distribution and proportion of different DOs in Corsican transitive sentences. In my survey, the informants chose *a*-marked DOs with human and/or referential referents as the 'best' option, more often then *a*-marked [indefinite], [inanimate] DOs.

Beyond the semantic accounts, I considered syntactic accounts on DOM (section 2.3). The approaches contained in these accounts can be aligned with two major ideas. Either *a*-marked DOs are analysed in terms of a dative construction (2.3.2), in which case the *a*-marking of (some) accusative objects equate to the *a*-marking of dative objects; this idea is motivated by the distributional pattern in ditransitive constructions, where dative objects are most consistently marked but *a*-marking is often omitted. Or all accusative objects are structurally the same, e.g. KPs or FPs, irrespective of any overt marking; some of them get *a*-marked due to certain feature constellations in the DP, but others do not because they do not fulfil the feature requirements. To understand better the structure of the DP, I dedicated section 2.3.1 to different split DP accounts, bearing in mind that neither the dative nor the accusative accounts provide a solution for the Corsican incompatibility of the definite article with the *a*-marker. These two items do not co-appear – or, in terms of my data, the co-appearance of both items was strongly disfavoured by the informants. The split accounts in section 2.3.1 are not strictly about DOM, but they do provide details about features, their distribution, and the structural consequences resulting from said feature distribution. They also consider case

assignment and semantic–interpretative aspects.

In Chapter 3 I have provided information about the Corsican language itself. As we see in section 3.2, Corsica was conquered and inhabited by peoples of different origin, shaping the appearance of the island in terms of architecture and landscape, but also affecting the language(s) spoken there. For the current state of Corsican, two periods may be emphasised in particular: when Pisa and Genoa reigned over Corsica from 1120 until 1768; and from 1768 until today, the political affiliation to France. These political power relations came along with linguistic power relations: under the Pisan and Genoese reign, Corsican was the spoken variety in a predominantly oral society; Tuscan served for written purposes, e.g. official documents and higher education. With the arrival of the French, Tuscan Italian was gradually superseded by French in the written domains, and only subsequent to that, did French conquered the oral Corsican domain(s). Nowadays French dominates in every domain of everyday life, from private relations to administration, politics and education. Even though Corsican has gained ground, e.g. the obligatory language class at university or the bilingual forms in public administration, French is nevertheless the dominant language.

With regard to intergenerational transfer, there is a noticeable gap. The French practised a strict language policy, prohibiting Corsican in public spheres, and in particular in school; children thus avoided speaking Corsican (to avoid punishment), and used French instead. In addition, France linked language competence of French to social success and advancement, by offering job opportunities in the French administration and military. The strategy of influencing the youngest, and with them also their families, caused Corsican to lose (at least) two generations of Corsican speakers. Only since the late 1970s have various initiatives been introduced – language courses, cultural associations, Corsican books, newspapers and news, etc. – to revitalise the language varieties, with varying degrees of success. While the status of Corsican as a regional language is politically safe and entitles Corsican to be represented in the public domain, the actual numbers of speakers vary greatly over time and with location as does the actual language competence of most Corsicans. On the one hand, the missing intergenerational transfer results in only a few speakers who consider Corsican their (only) mother tongue and show corresponding expertise; these consist more of the older generation than the youth. On the other hand, all Corsicans grew up in a language contact situation; French influences Corsican to different degrees.

The language behaviour of my informants may also reflect this situation. The occasional divergence between, for example two sentences of the same noun class, is probably due to variation between informants in terms of language competence, or their personal process of language

acquisition. I have not investigated the relationship between the answers given by informants, and language acquisition, i.e. the differences between Corsican as an L1 acquired in a mono- or bilingual setting, and Corsican as an L2.

In addition, diatopical differences may have influenced the informants' answers. Phonetically and lexically, the Corsican varieties show wide variation. Also, Corsican morphosyntax is probably less homogeneous than stated in the literature (see section 3.1.1 on diatopic differentiation). To investigate the particularities of single varieties, researchers seek informants who have lived continuously in the same place. But, as in other (European) countries and regions, there is a lot of migration in Corsica. Many people move due to education, work or personal circumstances, especially to the cities and urban centres, where language varieties mix and change. Socioculturally and administratively, this combines with the *polinomia* concept; it allows the co-existence and equivalence of all varieties spoken and written by Corsicans. Yet there are still discussions about which Corsican is 'the original' Corsican (see section 3.1.3).

Again, my group of informants did not allow for a systematic investigation of diatopical differences, due to limited numbers of representatives of single varieties. In fact, none of my informants lived without interruption in the place where s/he was born. With regard to the responses to my survey, the inconsistencies mentioned earlier (i.e. some sentences containing *a*-marked examples of a particular noun class were rated 'good', while other sentences containing *a*-marked examples of the same noun class not being rated 'good'), can probably attributed to distinctions between different language varieties, even though they may not be replicable due to the informants' personal migration history.

By providing information about the language history of Corsican, and relating the informants' backgrounds to the socio-historic context, I sought to show the interrelations between history and the informants' language behaviour. The investigation of relations or correlations between informants and language data, I leave for further research. For the present thesis, I have assumed all informants to be competent speakers of Corsican, and treated all their answers accordingly, i.e. included all rated and/or evaluated examples in the corpus.

Chapter 4 is dedicated to the state of the art of Corsican DOM. While the history of Corsican as an independent field of study is, in the German tradition of Romance linguistics, dated to the admission of Corsican as an independent entry in the *Lexikon der Romanischen Linguistik* (LRL) in 1988, Corsican DOM has been mentioned in the literature since the 19th century, although systematic investigations of the phenomenon have been scarce. In most cases, the literature mentions animacy

as the (only) DOM-trigger, and cites a few isolated examples – most often personal pronouns and proper names[human], and also (rarely) several different proforms. Chapter 4 gives an overview of: all noun classes considered in the literature (section 4.1); the role ascribed to the feature humanness (section 4.2); the incompatibility of Corsican DOM and the definite article (section 4.3); whether verb semantics influence the DOM-pattern in Corsican (section 4.4); how Corsican DOM can be analysed in theoretical terms, i.e. referring to disambiguation (section 4.5); and Giancarli's *Hypothèse de l'entonnoir* (section 4.6). This overview concludes with prescriptive rules for Corsican DOM found in textbooks and grammar books (section 4.7). This extensive review of the information provided in the literature revealed inconsistencies and gaps. These resulted in my research questions in section 4.8:

1. *How is humanness related to noun classes and syntactic structure?*

   Humanness seems to be a triggering factor for DOM. Nevertheless not all nominals[human] in DO-function get *a*-marked. For proforms, the *a*-marking corresponds quite well to the distinction between human/non-human DOs. But common nouns (with a definite article) are never a-marked, regardless of their animacy value. Proper nouns appear with and without the marking; names of persons are consistently *a*-marked, toponyms not always.

2. *Why do the definite article and the DOM-marker not co-appear?*

   In the literature, the definite article is said to *prevent* or *block* the realisation of the *a*-marker. This point of view suggests that, according to some mechanism, the definite article appears first and the *a*-marking then cannot take place. I investigated how this *blocking* process works, and to what extent other conditions determine this incompatibility.

3. *How to account for indefinite and negative quantifiers[human]?*

   Certain nominal items, such as indefinite and negative quantifiers[human], are difficult to reconcile with the traditional theories of Romance DOM. Being indefinites, their consistent *a*-marking contradicts common assumptions about DOM, its syntactic structure and semantic interpretation The projection of a DP is considered necessary to provide the syntactic structure for an additional functional layer related to case marking, i.e. KP/FP. So the question is, how can indefinite proforms trigger a DP? Further, in semantic terms, definiteness is a trigger for DOM – but indefinites cannot have any definite interpretation. How then do they trigger *a*-marking?

To answer these questions, I sought to base my investigation on a sound footing. I collected empirical language data to account for the descriptive gaps left in the literature, and to provide data for noun classes not considered in other studies, i.e. brand names or names of months. To this end, I prepared

a questionnaire and conducted two fieldtrips to Corsica. This is the topic of Chapter 5. From my prior experience with Corsican language data (cf. Neuburger 2008), I knew about the very low frequency of DOM in naturally spoken (or written) language. Hence, a specific questionnaire allowed for the elicitation of a higher quantity of relevant data. I decided to work with a combination of forced choice and evaluation tests. The forced choice initially asked the informant to choose, from a range of sentences differing only in terms of the left edge of the DP, the sentence s/he considered: the nominal appearing with or without DOM, and/or with or without the definite article, or even bare. Informants could subsequently also evaluate the option(s) they had not chosen as 'good'.

To compile examples for the questionnaire, I drew on sentences from the literature to test the examples cited therein. I added new sentences, checked by two native speakers, to complement the set of noun classes with those subclasses not represented in the literature (for details about the origin of the sentences, see Tables 60 and 61 in the appendix, Chapter 9). The compilation resulted in a questionnaire containing 50 examples, each of which offered a choice of at least two options.

I used this questionnaire for both of my fieldtrips, in June 2010 and September 2011, described in section 5.4. In advance of the fieldtrips, I emailed institutions, associations, and persons to make appointments, with little success. I managed to arrange only very few meetings in advance. On site, the difficulty of making contact with potential informants persisted. Persons who dedicate their work and passion to Corsican (e.g. teachers, librarians, editors, and authors) deliberately participated at the survey, and also introduced me to new informants. But I talked to hardly any unbiased Corsicans who were willing to participate. In the end, the answers of 46 informants were included in the corpus.

In Chapter 6 I provided the results of the questionnaire. Following the descriptive pattern used in Chapter , I divided the results according to the nominal classes to which the DOs being tested DOs belonged. Within these divisions I also took into account other aspects than nominal class, such as the distribution of definite articles and animacy. The impact of verb mode, which is relevant, for example, for Spanish DOM, but is not mentioned for Corsican, was not considered systematically. The only example with a mode-distinction between indicative and subjunctive did not result in remarkable differences. As a general result, we can conclude that Corsican DOM is not categorial: none of the classes provoked a 100% or 0% result; only single exponents of classes resulted in high or low numbers.

The noun classes mentioned in the literature (relating to Corsican DOM), and also used for the evaluation of the data from the QIab survey, are the same as those in the *referentiality* or *animacy* hierarchies presented in section 2.1.2. Within these hierarchies, the noun classes are hierarchically

ordered 'slots' which are determined by features values. The higher the animacy and/or referentiality values associated with one slot, the higher its position in the hierarchy. Nominals are associated with their respective nominal classes by means of the feature values they bear. Some of the hierarchy features, such as referentiality and definiteness, have a semantic interpretative effect, and are also reflected in a certain syntactic structure; for example, for referential and/or definite nominals we assume most commonly a DP-structure. But the nominal classes in the hierarchies do not correspond to one (or more) defined syntactic structure(s). Hence, phenomena which proceed along the lines of syntactic structure(s) are difficult to account for in terms of these hierarchies.

From the literature on DOM I knew that humanness of the DO has a strong impact on its marking. This is also reflected in the answers from my survey informants: *a*-marked kinship terms, prototypically referring to humans, like human names and any kind of proforms referring to humans, were preferably chosen by the informants as the 'good' option.

Proper nouns with inanimate referents, such as toponyms, brand names, or names of month and days, show a more diversified picture. Unmarked and *a*-marked toponyms were almost equally chosen as the best option. Toponyms where the *a*-marker and the definite article appeared simultaneously resulted in very low levels of acceptance. Brand names such as car names were only preferred with the definite article and without *a*-marking, while *a*-marking of a month name, *settembre*, used in a referential way, provoked many informants to choose it as the best option. The *a*-marking of names of days resulted in a similar pattern, but only a minority of informants chose *a*-marked year names as the best option.

Within the class of proforms, humanness is the crucial factor: *a*-marked personal pronouns, demonstratives[human], interrogatives[human], relatives[human], universal quantifiers[human], indefinite quantifiers[human], and negative quantifiers[human]; all of these were chosen as the best option by the majority of informants.

The survey data showed that humanness is a relevant trigger for DOM, but to a divergent degree between the different noun classes, with the selection of *a*-marked proforms resulting in the highest levels of acceptance in the dataset. Within the class of proper nouns, *a*-marking of human proper nouns resulted in higher acceptance levels than *a*-marking of inanimate proper nouns; within the class of common nouns, humanness did not seem to have any relevance at all, with *a*-marked common nouns rarely chosen as the best option, and common nouns marked by the definite article always preferred.

By comparing the *a*-marking of different nominal classes I have shown that humanness, as a trigger for *a*-marking, is delimited by the syntactic structure of the nominals; nominals[human] which

do not allow for a definite article, that is proforms and certain toponyms, get *a*-marked. Human common nouns bear a definite article, just as other referential common nouns do. So, the impact of humanness on triggering *a*-marking is delimited by syntax.

Looking at the syntactic structure of Corsican DOs, I focused on three issues: how, where, and why does overt case marking take place? Beyond humanness in semantic terms, López (2012) translated animacy into a morphosyntactic feature determining the projection of specific $DP_{[anim]}$: $D_{[anim]}$ is triggered by the animacy value of the nominal in question. By assuming such a mechanism, López accounts for the *a*-marking of indefinite and negative $quantifiers_{[human]}$. In terms of distribution, López's structure offers two functional heads, a D-head and a K-head. The case marking takes place within a KP-structure, where K hosts the *a*-marker, while the indefinite or negative $quantifier_{[human]}$ is located in DP. This structure not only accounts for the Spanish quantifiers López analysed, but also for their Corsican equivalents (see section 7.2.4).

Beyond these items, there remains the incompatibility of the definite article and the *a*-marker to approach, in distributional (and semantic–interpretative) terms. With regard to distribution, there are two options: we either assume a structure such as that from Rodríguez Mondoñedo's (2009); or such as that from López (2012) [KP [DP [NP …]]], which allows for two functional layers at the left edge. The projection of the DP can be triggered by inherent or ascribed feature values of the nominal, i.e. animacy, definiteness and/or referentiality. For Corsican, a [KP [DP [NP …]]]-structure would imply that DOs which allow for a definite article, e.g. [human], [referential] common nouns, project into a DP and appear with a definite article. A KP can be projected, but no *a*-marking takes places; *a*-marking will be prevented by some additional rule. If we assume that humanness and referentiality of the DO-nominal trigger *a*-marking, and DP is the requirement for the projection of a KP (and an overt realisation of /*a*/), definite common $nouns_{[human]}$ do fulfil the requirement to realise *a* in K. But *a*-marking does not take place. How can this constraint be formulated? Here things become (unnecessarily) complicated.

A simpler option is provided by Giusti's (2015) DP-structure with a single functional layer. The projection of the DP is driven by case requirements and, in terms of features, by an unvalued case feature [*u*Pred(Case)] hosted in D° and independent from referentiality requirements. The definite article is a case morpheme and merges into D°. Other features – such as referentially (i.e. the referential index ι), which is traditionally associated with D° or the D-elements, or [person] associated with the noun – are hosted in SpecDP. By separating referentiality from the D-element (i.e. the definite article), Giusti allows for the projection of a DP independently of referentiality features. For Corsican, this single function layer related to case can account for the complementary

distribution of the definite article and the *a*-marker; I assume that both items mark accusative case in Corsican. DOs with a high degree of animacy, i.e. humanness and/or referentiality, require overt case marking. The case requirement can then be fulfilled by either item (see section 7.2.5). The nominal itself determines which item fulfils the function: either it allows for a definite article, whereupon it is case marked by the article (common nouns); or it does not allow for definite articles, whereupon it is case marked by *a*.

This structure not only accounts for common nouns bearing a definite article and no *a*-marking, but can also account for the *a*-marking of indefinite and negative quantifiers[human]: the independent feature [*u*Case] is responsible for the projection of a DP, but not responsible for either definiteness or referentiality of the respective nominal. Hence, indefinite and negative quantifiers[human] project into a DP due to case requirements.

(At least) in Corsican, the advantage of a single functional layer responsible for case marking is obvious: it obviates the need to assume additional functional layers for items which do not provide the features to trigger these layers, such as indefinite and negative quantifiers[human] for definiteness. Also, a single syntactic position accounts easily for the complementary distribution of *a*-marking and the definite article: the D° hosts only one of the two items.

In summary, this dissertation offers a detailed insight into the distribution of DOM, its triggering factors, and its functions. To complement the picture of accusative case in Corsican that I have presented, a second survey is needed to include those nominals that are excluded from DOM due to their semantics. Then, we can ask whether all Corsican DOs require overt case marking – or whether DOs without a definite article or the *a*-marker actually remain caseless terms (as in Ormazabal & Romero's 2013), or are in some way incorporated into a *v*P-structure (cf. López 2012).

In fact, the compilation of an annotated corpus of naturally produced language may change the picture and offer new insights and possibilities for analysis. The data from the QIab survey could also be analysed considering the informants' language background; in terms of heritage speaker-studies Corsican could offer an interesting field of study. But also, in more traditional terms of language acquisition or the investigation of diatopical variation, the existing corpus may provide more insights.

These, and other issues of Corsican DOM and the study of Corsican morphosyntax, I leave for further research and current and future linguists of Corsican.

9 Appendix

This appendix contains all the documents mentioned in the main text with reference to the appendix. This includes the questionnaire used, but also more sophisticated representations and overviews of the data from the literature, and how they found their way into the questionnaire (Tables 60 and 61). It also includes more detailed, but anonymised, information about the informants (Tables 62 and 63).

Gracie per la Sua partecipazione!
Informazione generale: Questo questionario **NON È UN TEST!**

Prego il gentile informante /la gentile informatrice di
non usare libri, l'internet o altri mezzi per compilare il questionario
non collaborare con altre persone, **non** chieder aiuto à nessuno, **non** lavorare in gruppo
→ **È molto importante che le risposte sono autentiche e dimostrano la conoscenza e l'opinione di un singolo individuo.**

La prego di rimandare il questionario compilato à Kathrin Neuburger prima del
 15 ottobre 2011
- kneuburger@rom.uzh.ch
- Kathrin A. Neuburger; Karl-Marx-Str. 30; 12043 Berlin; Allemagne

Informazione personale:

Nome: _____

Natu in: _____

Annu: _____

Vissutu in cuntinente? Induve? Quantu tempu?: _____

Avale campa in: _____

Mistieru: _____

Lingua materna: _____

Altre lingue: _____

Lingua/e del padre/della madre: _____

Mistieru del padre/della madre: _____

Accetto che i dati saranno usati esclusivamente à scopo linguistico scentifico incl. la pubblicazione in forma di tesi dottorale.
Tutte i dati personali saranno anonimizzati: _____
                              Luogo, data   firma

1)   *a. U sgiò Marcelli chjama à so fratellu:*
          b. U sgiò Marcelli chjama so fratellu:
          c. U sgiò Marcelli chjama u so fratellu:
          d. _____

2)   *a. Mandi à qualchissia à piglià issa cosa?*
          b. Mandi qualchissia à piglià issa cosa?
          c. _____

3)   *a. Mandi à qualcunu per aiutammi?*
          b. Mandi qualcunu per aiutammi?
          c. _____

---

4)   *a. Petru hà avutu un'accidente cù a vittura. Hà dannighjatu à Fiat nova d'Anna.*
          b. Petru hà avutu un'accidente cù a vittura. Hà dannighjatu a Fiat nova d'Anna.
          c. _____

5)   *a. Cunnosci à unu chì sà riparà a mo vittura?*
          b. Cunnosci unu chì sà riparà a mo vittura?
          c. _____

          d. Cunnosci à unu chì sappii riparà a mo vittura?
          e. Cunnosci unu chì sappii riparà a mo vittura?
          f. _____

---

6)   Dui omi anu arrubatu i soldi di a signora Muralli.
         a. À qual'aghju da seguità? Aghju da seguità à quistu o à quillu?
         b. Qual'aghju da seguità? Aghju da seguità quistu o quillu?
         c. _____

         d. À chi devu seguità?
         e. À quale devu seguità?
         f. _____

---

7)   Petru stà à u telefunu. Dumanda Maria:
         a. Quale hè chì hà chjamatu? – Hà chjamatu Ghjuvanni, vole vene.
         b. À quale ai chjamatu? – Aghju chjamatu à Ghjiseppu
         c. _____

---

8)   Anghjula hè turnata in casa.
         a. À quale hè ch'hè vinutu oghje à u scagnu ? - Hè vinutu u sgiò Baldoni.
         b. Chì hè ch'hè vinutu oghje à u scagnu ? - Hè vinutu u sgiò Baldoni.
         c. Quale hè ch'hè vinutu oghje à u scagnu ? - Hè vinutu u sgiò Baldoni.
         d. _____

9)   *a. È eiu aghju vistu à to parente u duttore*
         b. È eiu aghju vistu u to parente u duttore
         c. _____

---

10)   Petru si vole marità. Ma u so amicu Marcu hè assai indecisu:
         a. Ùn hè micca difficiule à tene à moglie fidele in issi tempi?
         b. Ùn he micca difficiule à tene moglie fidele in issi tempi?
         c. _____

11)   *a. Aspittaremu à 2011 per u nostru matrimoniu.*
         b. Aspittaremu 2011 per u nostru matrimoniu.
         c. Aspittaremu u 2011 per u nostru matrimoniu.
         d. _____

12)   Chì mese avete sceltu per a festa?
         a. Avemu sceltu à settembre
         b. Avemu sceltu u settembre
         c. _____

---

13) In chjesa:
    a. Prighemu à a madonna
    b. Prighemu a madonna
    c. _____
    d. A pregu per a salute di u mo babbu
    e. Li pregu per a salute di u mo babbu
    f. _____

14)   *a. Prighemu à Ghjesu Cristu*
    b. Prighemu Ghjesu Cristu
    c. _____

    d. U pregu per a salute di u mo babbu
    e. Li pregu per a salute di u mo babbu
    f. _____

---

15)   *a. pregà à qualcunu per un favore*
    b. pregà qualcunu per un favore
    c. _____

16)   *a. pregà à Diu - dì una preghera*
    b. pregà Diu - dì una preghera
    c. _____

---

17) In piazza:
    a. Cunnoscu à Battistu ma un cerc'à ellu. Cercu à Maria.
    b. Cunnoscu Battistu ma un cercu ellu. Cercu Maria.
    c. _____

18)   *a. Eiu invece cercu à a mo Peugeot - l'ingareraghju culà.*
    b. Eiu invece cercu a mo Peugeot - l'ingareraghju culà.
    c. _____

---

19) Francescu hè riturnatu in casa. Hà fattu un viaghju in Allimagna è u conta à u so amicu Paulu:
    a. F: Cunnoscu à Berlinu.
    b. F: Cunnoscu Berlinu.
    c. P: Sulamente à Berlinu?
    d. P: Sulamente Berlinu?
    e. F: Iè, à Berlinu, u cunnoscu, ma l'Allimagna ùn a cunnoscu micca.
    f. F: Iè, Berlinu, u cunnoscu, ma l'Allimagna ùn a cunnoscu micca.
    g. _____
    _____

---

20)   *a. Aspettu à luni, ghjè a festa di Anna.*
          b. Aspettu luni, ghjè a festa di Anna.
          c. Aspettu u luni, ghjè a festa di Anna.
          d. _____

21)   Vedi culà, ghjè Francescu!
          a. Ùn vecu à nimu
          b. Ùn vecu nimu
          c. _____

22)   *a. Ùn vecu à nisunu*
          b. Ùn vecu nisunu
          c. _____

23)   *a. Vecu à qualchidunu, ma ùn hè micca Francescu, hè Ghjiseppu.*
          b. Vecu qualchidunu, ma ùn hè micca Francescu, hè Ghjiseppu.
          c. _____

          d. Vecu à qualchidunu.
          e. Vecu qualchissia.
          f. Vecu qualchidunu.
          g. Vecu à qualchissia.
          h. _____

24)   *a. Anna hà cumpratu à a Fiat nova*
          b. Anna hà cumpratu à issa Fiat nova
          c. Anna hà cumpratu a Fiat nova
          d. Anna hà cumpratu issa Fiat nova
          e. _____

25)   Vedi culà, ghjè a mo vittura nova!
          a. Ùn vecu à nulla
          b. Ùn vecu nulla
          c. _____

26)   Eri hà fattu un tempacciu!
          a. U ventu forte hà quasi lampatu à babbu.
          b. U ventu forte hà quasi lampatu babbu.
          c. U ventu forte hà quasi lampatu babbu in terra.
          d. U ventu forte hà quasi lampatu à babbu in terra.
          e. _____

27)   *a. U ventu hà quasi lampatu à mo Peugeot*
          b. U ventu hà quasi lampatu a mo Peugeot
          c. _____

231

28)   a. *À Bastia, ùn l'aghju mai vista prima. Devu andacci.*
          b. *Bastia, ùn l'aghju mai vista prima. Devu andacci.*
          c. _____

29)   a. *Stà à sente à mammata… ma, a stai à sente?*
          b. *Stà à sente mammata… ma, a stai à sente?*
          c. _____

          d. *Stà à sente babbitu… ma, u stai à sente?*
          e. *Stà à sente à babbitu… ma, u stai à sente?*
          f. _____

30)   In magazinu*:*
          a. *Aghju compratu à Persil - hè in offerta è u più ecunomicu detergente chi ci hè.*
          b. *Aghju compratu u Persil - hè in offerta è u più ecunomicu detergente chi ci hè.*
          c. *Aghju compratu Persil - hè in offerta è u più ecunomicu detergente chi ci hè.*
          d. _____

31)   Chì faci?
          a. *Guardu à Lione, u cane di Petru: hà manghjatu un scarpu.*
          b. *Guardu Lione, u cane di Petru: hà manghjatu un scarpu.*
          c. _____

32)   Eri ci hè statu un grande focu.
          a. *U fume cupria à Portivechju*
          b. *U fume cupria Portivechju*
          c. _____

33)   a. *U fume copre à La Marana*
          b. *U fume copre a Marana*
          c. _____

34)   I dui vanu à cavalcà:
          a. *Petru s'hè coltu à U Sciroccu è Maria s'hè colta à Bella.*
          b. *Petru s'hè coltu U Sciroccu è Maria s'hè colta à Bella.*
          c. *Petru s'hè coltu U Sciroccu è Maria s'hè colta Bella.*
          d. _____

35)   *Chì faci?*
          a. *Leghju à Goethe.*
          b. *Leghju Goethe.*
          c. _____

36) *a. Dò i soldi à i poveri - per aiutà à chì ne hà bisognu*
     b. Dò i soldi à i poveri - per aiutà chì ne hà bisognu
     c. _____

37) *a. Ci vole à fà ghjurà à tutti chì elli seranu d'accunsentu cun tè*
     b. Ci vole à fà ghjurà tutti chì elli seranu d'accunsentu cun tè
     c. _____

38) *a. Devu ringrazià tutti quelli chì mi anu aiutatu.*
     b. Devu ringrazià à tutti quelli chì mi anu aiutatu.
     c. _____

39) *a. Chjamu à ognunu à tene u rollu chì hè u soiu*
     b. Chjamu ognunu à tene u rollu chì hè u soiu
     c. _____

40) *U mo cane ghjè nant'à strada, si batte cù u cane di u vagabondu.*
     a. Petru hà tumbatu à unu senza ferisce l'altru.
     b. Petru hà tumbatu unu senza ferisce l'altru.
     c. _____

41) *a. Fate puru sapè à tuttu u mondu chì sò qui.*
     b. Fate puru sapè tuttu u mondu chì sò qui.
     c. _____

42) *a. E cundizioni ughjinche attiranu i spiculatori è a maffia*
     b. E cundizioni ughjinche attiranu à i spiculatori è à a maffia
     c. E cundizioni ughjinche attiranu i spiculatori è à maffia
     d. _____

43) *a. Ùn vegu à nisciun'omu.*
     b. Ùn vegu imu.
     c. _____

44) *a. Cunnoscu à iss'omu.*
     b. Cunnoscu iss'omu.
     c. _____

45)   *a. U focu distrughje Corti.*
           b. U focu distrughje à Corti.
           c. _____

46)   *a. Ùn ai micca vistu à babbu? Hè esciutu 5 minuti fà.*
           b. Ùn ai micca vistu babbu? Hè esciutu 5 minuti fà.
           c. _____

47)   *a. Petru hà avutu un accidente cù a vittura.*
           b. Hà avutu un accidente cù a vittura, Petru.
           c. _____

48)   *a. Maria hà vintu u loto.*
           b. Hà vintu u loto, Maria.
           c. _____

49)   *a. Hà chjamatu à Petru.*
           b. Hà chjamatu Petru.
           c. _____

50)   *a. Devu ringrazià tutti.*
           b. Devu ringrazià à tutti.
           c. _____

Osservazioni:
_____
_____
_____
_____
_____
_____
_____

Table 61: QIab – Examples inspired from the literature; taken over in original or modified[75] *form (*Chronologically ordered by bibliographical reference, i.e. source): **Obligatory DOM**

Comment/source	Original example	no.[76]	Example in my questionnaire QIab[77]
"[…] la preposizione precede o un **pronome** o un **nome proprio**; in tutti gli altri casi la preposizione manca […]" (Bottiglioni 1957c: 121)			
(Bottiglioni 1933–1942: map. 1860)	*a ccual' agghiu da siguidà?*	Iab.06	*(À/ø) qual'aghju da seguità?*
		Iab.06	*À (quale/chi) devu seguità?*
(Bottiglioni 1933–1942: map 228)	*belude gh'e gi voga o ghe gi mandi a ccalchissia?*	Iab.02	*Mandi (à/ø) qualchissia à piglia issa cosa?*
		Iab.03	*Mandi (à/ø) qualcunu per aiutammi?*
(Bottiglioni 1933–1942: map. 1778)	*prighemm' a gghiesu gristu e a a madonna*	Iab.13	*Prighemu (à) a Madonna*
		Iab.14	*Prighemu (à/ø) Ghjesu Cristu.*
(Bottiglioni 1933–1942: map. 1900)	*cunnoscu a bbattistu ma un cerc'ad ellu*	Iab.17	*Cunnoscu (à/ø) Battistu ma un cerc'(à/ø) ellu. Cercu (à/ø) Maria.*
"[…] la construction du 'complément d'objet direct' 1) avec la preposition *à* quand il s'agit d'un **nom propre, non précédé de l'article** […] Le **nom de lieux** se distribuent de la même façon, selon qu'ils ont ou n'ont pas l'article […" (Marcellesi 1983: 403)			
"Les 'noms propres' non prédéterminés et prénoms de presonnes […], d'animaux, de localité […]" (Marcellesi 1986: 131)			

---

[75] Modification was predominantly undertaken through changing vocabulary and adding a wider context.

[76] There reference numbers correspond to my system of questionnaire-numbering.

[77] Elements in parentheses correspond to the different varieties of sentences; i.e. *(À/ø) qual' aghju da seguità? = À qual' aghju da seguità? / Quale aghju da seguità?*

(Marcellesi 1986: 137; ex. 7)	*U fumu copri à Portuvechju*	Iab.32	*U fume cupria (à/ø) Portivecchju.*
		Iab.45	*U focu distrughje (à/ø) Corti.*
		Iab.19	*Cunnoscu (à/ø) Berlinu.*
		Iab.28	*(À/ø) Bastia, ùn l'aghju mai vista prima. Devu andacci.*
(Marcellesi 1986: 137; ex. 3)	*Vidimu U Conti* [village-name]	Iab.33	*U fume copre (à/ø) La Marana*
(Marcellesi 1986: 137; ex. 4)	*Ha purtatu U Sciroccu*	Iab.34	*Petru s'hè soltu (à/ø) U Sciroccu è Maria s'hè coltu (à/ø) Bella.*
(Marcellesi 1986: 137; ex. 9)	*U ventu hà guasgi lampatu à babbu.*	Iab.26	*U ventu hà guasgi lampatu (à/ø) babbu.*
(Marcellesi 1986: 137; ex. 14)	*Temu à Diu.*	Iab.16	*Pregà (à/ø) Diu – dì una preghera.*
"Désormais la règle est que l'on doit avoir *à* caduc quand le référé est humain" (Marcellesi 1986: 133)			
(Marcellesi 1986: 138; ex. 32)	*Vigu à calchiadunu, a tutti.*	Iab.23	*Vecu (à/ø) qualchidunu.*
(Marcellesi 1986: 138; ex. 33)	*Un vigu à nimu, a nisciunu.*	Iab.21	*Ùn vecu (à/ø) nimu.*
		Iab.22	*Ùn vecu (à|ø) nisunu.*
(Marcellesi 1986: 138; ex. 33)	*Un vigu nisciunu omu.*	Iab.43	*Ùn vegu (à/ø) nisciun'omu.*
(Marcellesi 1986: 138; ex. 34)	*Vigu quiss'omu.*	Iab.44	*Cunnoscu (à/ø) iss'omu.*
"Parmi les règles complexes qui déterminent l'apparition de *a* en corse, il convient de mettre en valeur l'opposition entre noms autodéterminés et noms non autodéterminés, qui semble plus appropriée que l'opposition humain/non humain" (Chiorboli 1987: 87)	*ha chiamatu a Petru*	Iab.07	*À quale ai ghjamatu? Aghju chjamatu à Ghjuseppu.*

"The prepositional accusative, used when the object of the verb is a **proper name**, a **disjunctive personal pronoun** or one of a set of **kinship terms** […]" (Thiers 1993: 255)	*Vecu à ellu, a Petru, a mammata*	Iab.17	*Cunnoscu (à/ø) Battistu ma ùn cerc' (à/ø) ellu. Cercu (à/ø) Maria.*
"[…] l'oggetto diretto introdotto dalla preposizione *a* davanti [a nome proprio o] a **pronome indefinito** […]" (Nesi 2002: 963)	*chjama a Marianghjula chiama a calchissia*	Iab.02	*Mandi (à/ø) qualchissia à piglià issa cosa?*
		Iab.03	*Mandi (à/ø) qualcunu per aiutammi?*
		Iab.05	*Cunnosci (à/ø) unu chì sà riparà a mo vitura?*
		Iab.15	*Pregà (à/ø) qualcunu per un favore?*
		Iab.23	*Vecu (à/ø) qualchidunu, ma ùn hè micca Francescu, hè Ghjuseppu.*
		Iab.40	*Petru hà tumbatu (à/ø) unu senza ferisce l'altru.*
(Giacomo-Marcellesi 1997: 41, 42; ex. 103)	*U focu a varcat'à Tavignanu!*	Iab.32	*U fumu copria (à/ø) Portuvecchju.*
		Iab.33	*U fumu copre à La Marana*
		Iab.33	*U fumu copre a Marana.*
		Iab.45	*U focu distrughe (à/ø) Corti*
(Giacomo-Marcellesi 1997: 41, 42; ex. 105)	*Un vig'à nimu.*	Iab.21	*Ùn vecu (à/ø) nimu.*
		Iab.22	*Ùn vecu (à/ø) nisunu.*
(Casta 2003:119)	*Aghju visto à to parente u duttore.*	Iab.09	*È eu aghju vistu (à/u) to parente u duttore.*
"Devant les **millésimes (sans article)**, les **noms de mois** ou de **jours**, traités en corse comme des noms propres COD […]" (Casta 2003: 120)	*Aspittaremu à 2005.*	Iab.11	*Aspittaremu (à/ø/u) 2011 per u nostru matrimoniu.*
	Aspittaremu à nuvembre.	Iab.12	*Avemu sceltu (à/u) settembre.*
	Aspittaremu à luni.	Iab.20	*Aspittu (à/ø/u) luni, ghjè a festa di Anna.*
(Manzini & Savoia 2006: 513 [vol. II]; ex. 251)	*Ha chjamatu à so fratelli.[78]*	Iab.01	*U sciò Marcelli chjama (à/ø/u) so fratellu.*

---

[78] Example given (from Manzini & Savoia 2006) has been transcribed from IPA to Corsican orthography for convenience.

Table 62: Examples inspired by the descriptions of/comments in the literature on reasons to **prevent Corsican DOM**

Comment/source	no.	Examples
"[…] la construction du 'complément d'objet direct' […] 2) sans préposition quand c'est un **nom commun précédé de l'article**" (Marcellesi 1983: 403, fn.17)	Iab.42	*E cundizioni aghjinche attiranu (à/ø) i spiculatori è (à/ø) a maffia.*
"Les 'noms propres' quand ils sont constitués d'une suite 'Art.+Nom' interdisent le maintien de *à* caduc […]" (Marcellesi 1986: 131)		
"On notera que le corse pratique […] l'acccusatif prepositionnel […]: **l'article bloque d'ordinaire l'apparition de la préposition** […]" (Chiorboli 1991a: 77)	Iab.13	*Prighemu (à/ø) a Madonna.*
	Iab.33	*U fume core (à/ø) La Marana.*
	Iab.34	*Petru s'hè soltu (à/ø) U Sciroccu è Maria s'hè coltu (à/ø) Bella.*

Table 63: Overview informants, fieldwork June 2010

Code	Current residence	Born in the same region?	Years abroad	Bilingual Corsican-French	Bilingual French-Corsican	L2 Corsican
Aged 18–35 (born between 1975 and 1992)						
m82f.Ia	Bastia	yes	0			1
Aged 36–65 (born between 1945 and 1974)						
m71f.Ia	Bastia	yes	1			1
m69cf.Ia	Aiacciu	no	21	1		
m62fc.IaIbII	Bastia	yes	10		1	
f59fc.Ia	Bastia	yes	15		1	
m57cf.Ia	Bastia	no	6	1		
f56f.Ia	Aiacciu	no	15			1
m53cf.Ia	Bastia	yes	25	1		
m52cf.Ia	Bastia	no	20	1		
m51cf.IaIb	Aiacciu	yes	10	1		
m47cf.Ia	Bastia	yes	9	1		
Aged 66+ (born 1944 and before)						
m31cf.IaIbII	Balagna	no	20	1		

Table 64: Overview informants, fieldwork September 2011

Code	Current residence	Born in the same region?	Years abroad?	L1 Corsican	Bilingual Corsican-French	Bilingual French Corsican	L2 Corsican
Aged 18–36 (born between 1975 and 1993 [1994[79]])							
m94c.IbII	Corti	no	0	1			
f94f.IbII	Bastia	?	0				1
f93c.IbII	Bastia	?	0	1			
f93f.IbII	Ile-Rousse	no	0				1
f93cf.IbII	Corti	no	0		1		
f93fc.II	?	?	?			1	
m93cf.IbII	Corti	no	0		1		
f92fc.II	?	?	?			1	
f92f.IbII	Sermanu	no	0				1
m92c.IbII	Venzulasca	yes	0	1			
m92cf.IbII	Corti	no	0		1		
m91cf.IbII	Corti	no	0		1		
f91f.IbII	Corti	no	0				1
m83f.IbII	Corti	?	2				1
m81f.IbII	Aiacciu	yes	6				1
f79f.IbII	Corti	no	13				1
m77cf.IbII	Bastia	yes	11		1		
m76f.Ib	Cervioni	no	0				1
m75cf.IbII	Corti	no	1		1	1	

---

[79] The age-range has been widened as a result of the participation of two studentens (aged 17) in the survey conducted in class.

Table 65 (cont'd): Overview informants, fieldwork September 2011

Code	Current residence	Born in the same region?	Years abroad?	L1 Corsican	Bilingual Corsican-French	Bilingual French Corsican	L2 Corsican
Aged 37–66 (born between 1945 and 1974)							
m65c.IbII	Aiacciu	no	1	1			
m64cf.IbII	Francardu	yes	0		1		
f62f.IbII	Cervioni	no	0				1
m62fc.IaIbII	Bastia	yes	10			1	
f61fc.IbII	Bastia	no	0			1	
m60f.IbII	Bucugnanu	no	13				1
f56c.IbII	Bastia	yes	0	1			
m56f.IbII	San Ghjuliani	no	06				1
f55c.IbII	?	?	0	1			
m51cf.IaIb	Aiacciu	yes	10		1		
m46c.IbII	U Pedigrighju	yes	7	1			
f46f.Ib	Cervioni	no	20				1
m45cf.IbII	Biguglia	yes	02		1		
Aged 66+ (born 1944 and before)							
m43cf.IbII	Bastia, Piedicroce	yes	25		1		
f39f.IbII	Bastia	yes	14				1
m38c.IbII	?	no	40	1			
m31cf.IaIbII	Balagna	no	20		1		
m25cf.Ib	Bastia	yes	11		1		

All tables from Chapter 6 with percentages. The numbering refers of the numbering in Chapter 6. The distribution of all tokens.

1. Kinship terms

Table 18.%: Kinship terms

	total	*good*	*bad*	?	ø
à + kinship term	84 100%	68 **80.95%**	2 2.38%	0 0%	14 16.67%
ø + kinship term	84 100%	16 19.05%	18 21.43%	1 1.19%	49 **58.33%**
à + possessive + kinship term	179 100%	84 **46.93%**	21 11.73%	3 1.68%	71 **39.66%**
ø + possessive + kinship term	133 100%	21 15.79%	30 22.56%	1 0.75%	81 **60.90%**
ø + def.art. + poss. + kinship term	93 100%	78 **83.87%**	6 6.45%	2 2.15%	7 7.53%

Table 19.%: Kinship term babbu 'dad'

(à/ø)+ kinship term 'babbu'		total	*good*	*bad*	?	ø
Iab.46a	Ùn ai micca vistu à babbu?	41 100%	34 82.93%	0	0	7 17.01%
Iab.46b	Ùn ai micca vistu babbu?	41 100%	9 21.95%	9 21.95%	0	23 56.1%

Table 20.%: Kinship term fratellu 'brother'

(à/ø/def.art.) + poss. + kinship term		total	*good*	*bad*	?	ø
Iab.01a	*U sgiò Marcelli chjama à so fratellu*	46 100%	4 8.7%	10 21.74%	1 2.17%	31 67.39%
Iab.01b	*U sgiò Marcelli chjama so fratellu*	46 100%	1 2.17%	12 26.09%	1 2.17%	32 69.57%
Iab.01c	*U sgiò Marcelli chjama u so fratellu*	47 100%	40 85.11%	3 6.38%	1 2.13%	3 6.38%

2. Proper nouns

Table 21.%: Ghjesu Cristu and *A Madonna*

'We worship …'		total	*good*	*bad*	?	ø
Iab.13a	*Prighemu à a Madonna.*	46 100%	9 **19.57%**	8 **17.39%**	2 4.35%	27 **58.70%**
Iab.13b	*Prighemu ø a[DET] Madonna*	46 100%	37 **80.43%**	2 4.35%	2 4.35%	5 **10.87%**
Iab.13c/ alternative	*Prighemu à (*a[DET]) Madonna*	46 100%	1 2.14%			
Iab.14a	*Prighemu à Ghjesu Cristu*	46 100%	40 **86.96%**	1 2.17%	0 0%	5 **10.87%**
Iab.14b	*Prighemu ø Ghjesu Cristu*	45 100%	10 **22.22%**	7 5.56%	2 4.44%	26 **57.78%**

Table 22.%: Battistu [male person]

'I know Battistu…'		total	*good*	*bad*	?	ø
Iab.17a	*Cunnoscu à Battistu…*	46 100%	40 **86.96%**	0 0	0 0	6 13.04%
Iab.17b	*Cunnoscu ø Battistu…*	48 100%	8 16.67%	11 22.92%	2 2.08%	28 **58.33%**

Table 23.%: Goethe [metonymically used to refer to his book(s)]

'I read Goethe'		total	*good*	*bad*	?	ø
Iab.35a	*Leghju à Goethe*	46 100%	36 **78.26%**	2 4.35%	1 2.17%	7 15.22%
Iab.35b	*Leghju ø Goethe*	45 100%	13 **28.89%**	8 17.78%	0 0	52 53.33%

Table 24.%: Lione [dog name]

'I watch Lione, the dog…'		total	*good*	*bad*	?	ø
Iab.31a	*Guradu à Lione, u cane…*	45 100%	34 **75.56%**	2 4.44%	0	9 20.00%
Iab.31b	*Guradu ø Lione, u cane…*	45 100%	22 **24.44%**	8 **17.78%**	2 4.44%	24 **53.33%**

Table 25.%: U Sciroccu [horse name]

'Petru has chosen *U Sciroccu…*'		total	*good*	*bad*	?	ø
Iab.34a	*Petru s'hè coltu à U Sciroccu*	46 100%	20 **43.48%**	4 8.70%	1 2.17%	21 **45.65%**
Iab.34b	*Petru s'hè coltu ø U Sciroccu*	46 100%	7 **15.22%**	11 **23.91%**	0 0	28 **60.87%**
Iab.34/ alternative	*Petru s'hè coltu à (*U) Sciroccu*	46 100%	9 **19.57%**			

Table 26.%: Proper nouns [inanimate]

	total	*good*	*bad*	?	ø
à	731 100%	249 **34.09%**	119 16.28%	24 3.28%	338 **46.24%**
ø	867 100%	495 **57.09%**	62 7.15%	39 4.50%	271 **31.26%**

Table 27.%: settembre 'September'

'We have chosen September'		total	*good*	*bad*	?	ø
Iab.12a	*Avemu sceltu à settembre.*	46 100%	32 **69.57%**	7 15.22%	2 4.35%	5 **10.87%**
Iab.12b	*Avemu sceltu u settembre.*	46 100%	7 **15.22%**	10 21.74%	1 2.17%	28 **60.87%**
Iab.12c/ alternative	*Avemu sceltu ø settembre*	46 100%	11 **23.91%**			

Table 28.%: luni 'Monday'

'I wait for Monday…'		total	*good*	*bad*	?	ø
Iab.20a	*Aspettu à luni…*	46 100%	25 **54.35%**	6 13.04%	1 2.17%	14 **30.43%**
Iab.20c	*Aspettu u luni…*	46 100%	12 **26.09%**	5 10.87%	2 4.35%	27 **58.70%**
Iab.29b	*Aspettu ø luni…*	46 100%	25 **54.35%**	1 2.17%	2 4.35%	18 **39.13%**

244

Table 29.%: 2011 '[the year] 2011'

'We wait for 2011…'		total	*good*	*bad*	?	ø
Iab.11a	*Aspittaremu à 2011…*	46 100%	13 **28.26%**	8 17.39%	3 6.52%	22 **47.83%**
Iab.11c	*Aspittaremu u 2011…*	46 100%	25 **54.35%**	1 2.17%	4 8.70%	16 **34.78%**
Iab.11b	*Aspittaremu ø 2011…*	46 100%	13 **28.26%**	3 6.52%	5 10.87%	25 **54.35%**

Table 30.%: Fiat

'…bought the/this new Fiat'		total	*good*	*bad*	?	ø
Iab.24a	*… cumpratu à a*[DET] *Fiat nova.*	46 100%	0 0	12 **26.09%**	2 4.35	32 **69.57%**
Iab.24b	*… cumpratu à issa*[DEM] *Fiat nova.*	46 100%	1 **2.17%**	11 23.91%	1 2.17%	33 **71.74%**
Iab.24c	*… cumpratu ø a*[DET] *Fiat nova.*	46 100%	41 **91.11%**	0 0	2 4.44%	2 **4.44%**
Iab.24d	*… cumpratu ø issa*[DEM] *Fiat nova.*	46 100%	39 **84.78%**	0 0	0 0	7 **15.22%**
'… damaged the new Fiat of Anna'						
Iab.4a	*… dannighjatu à Fiat nova di Anna.*	46 100%	4 **8.70%**	12 26.09%	1 2.17%	29 **63.04%**
Iab.4b	*… dannighjatu a*[DET] *Fiat nova di Anna.*	46 100%	40 **86.96%**	1 2.17%	0 0	5 **10.87%**

Table 31.%: mo Peugeot

'looking for my Peugeot…'		total	*good*	*bad*	?	ø
Iab.18a	*…cercu à a*[DET] *mo Peugeot…*	46 100%	3 **6.52%**	**9** **19.57%**	2 4.35%	32 **69.57%**
Iab.18b	*…cercu a*[DET] *mo Peugeot…*	46 100%	40 **86.96%**	1 2.17%	2 4.35%	3 **6.52%**
'…overturn my Peugeot'						
Iab.27a	*… lampatu à mo Peugeot.*	45 100%	6 **13.33%**	11 24.44%	1 2.22%	27 **60.00%**
Iab.27b	*… lampatu a*[DET] *mo Peugeot.*	45 100%	36 **80.00%**	2 4.44%	2 4.44%	5 **11.11%**

Table 32.%: Persil [name of washing powder]

'I have bought Persil…'		total	*good*	*bad*	?	ø
Iab.30a	*...cumpratu à Persil...*	46 100%	12 **26.09%**	10 21.74%	1 2.17%	23 **50.00%**
Iab.30b	*...cumpratu u[DET] Persil...*	46 100%	28 **60.87%**	3 6.52%	1 2.17%	14 **30.43%**
Iab.30c	*...cumpratu ø Persil...*	46 100%	12 **26.09%**	9 19.57%	2 4.35%	23 **50.00%**

Table 33.%: Toponyms

	total	*good*	*bad*	?	ø
à + toponym	272 100%	149 **54.78%**	27 9.93%	9 3.31%	87 **31.99%**
ø + toponym	271 100%	136 **50.17%**	24 8.86%	15 5.54%	96 **35.42%**

Table 34.%: (L)A Marana

'The smoke covers *A Marana*'		total	*good*	*bad*	?	ø
Iab.33a	*... copre à La[DET] Marana.*	46 100%	4 **8.70%**	6 13.04%	2 4.35%	34 **73.91%**
Iab.13a	*... copre A[DET] Marana.*	46 100%	41 **89.13%**	2 4.35%	1 2.17%	2 **4.35%**

3. Common noun

Table 35.%: Common nouns [human]

	total	*good*	*bad*	?	ø
à + def.art.	90 100%	11 12.22%	22 **24.44%**	1 1.11%	56 **62.22%**
ø + def.art.	45 100%	39 **86.67**	0 0	0 0	6 13.33%
à + demonstrative	41 100%	11 26.83%	9 **21.95%**	0 0	21 **51.22%**
ø + demonstrative	41 100%	28 **68.29%**	2 4.88%	2 4.88%	9 21.95%
à + negative quantifier	45 100%	15 **33.33%**	7 15.38%	1 2.22%	22 **48.49%**
ø + negative quantifier	45 100%	18 **40.00%**	5 11.11%	1 2.22%	21 **46.67%**

Examples for common nouns

		total	*good*	*bad*	?	ø
Iab.42a	*E cundizioni ughjinche attiranu i spiculatori*	45 100%	39 89.67%	0	0	6 13.33%
Iab.42b	*E cundizioni ughjinche attiranu à i spiculatori*	45 100%	6 13.33%	10 22.22%	1 2.22%	28 62.22%
Iab.44a	*Cunnoscu à iss'omu*	41 100%	11 26.83%	9 21.95%	0	21 51.22%
Iab.44b	*Cunnoscu iss'omu*	41 100%	28 68.29%	2 4.88%	2 4.88%	9 21.95%
Iab.43a	*Ùn vegu à nisciun'omu*	45 100%	15 33.33%	7 15.38%	1 2.22%	22 48.49%
Iab.43b	*Ùn vegu nisciun'omu*	45 100%	18 40.00%	5 11.11%	1 2.22%	21 46.67%

4. Proforms

Table 36.%: Personal pronouns

	total	*good*	*bad*	?	ø
à + personal pronoun	46 100%	40 **86.96%**	0 0	6 13.04%	0 0
ø + personal pronoun	48 100%	8 16.67%	11 22.92%	28 **58.33%**	1 2.08%

Table 37.%: Demonstratives *quistu, quillu* [human]

'Have I to follow this or that one?'		total	*good*	*bad*	?	ø
Iab.6a	*Adhju da seguità à quistu o à quillu?*	45 100%	37 **82.22%**	2 4.44%	0 0d	6 13.33%
Iab.6b	*Adhju da seguità ø quistu o ø quillu?*	45 100%	6 13.33%	7 15.56%	6 6.67%	29 **64.44%**

Table 38.%: Relative pronoun [human]: *chì*

'I give money to the poor, to help whoever needs it'		total	*good*	*bad*	?	ø
Iab.36a	*Dò i soldi à i poveri, per aiutà à chì ne ha bisognu.*	44 100%	37 **84.09%**	3 6.82%	0 0	4 9.09%
Iab.36b	*Dò i soldi à i poveri, per aiutà ø chì ne ha bisognu.*	44 100%	8 18.18%	6 13.64%	1 2.27%	29 **65.91%**

Table 39.%: Interrogatives *à quale – chì – quale*

'Who came to the office today? …'		total	*good*	*bad*	*?*	*ø*
Iab.8a	*À quale hè ch'hè vinutu oghje à u scagnu?*	46 100%	1 2.17%	13 **28.26%**	0 0	32 **69.57%**
Iab.8b	*Chì hè ch'hè vinutu oghje à u scagnu?*	46 100%	2 4.35%	11 **23.91%**	1 2.17%	32 **69.57%**
Iab.8c	*Quale hè ch'hè vinutu oghje à u scagnu?...*	46 100%	44 **95.65%**	1 2.17%	0 0	1 2.17%

Table 40.%: à quale [human] in DO-function

	total	*good*	*bad*	*?*	*ø*
à quale	135 100%	115 **85.19%**	7 5.19%	1 0.74%	12 8,89%
ø quale	45 100%	6 **13.33%**	7 15.56%	3 6.67%	29 64.44%

Table 41.%: Universal quantifiers [human]

	total	*good*	*bad*	*?*	*ø*
à + universal quantifier	120 100%	99 **82.50%**	4 3.33%	3 2.50%	14 11.67%
ø + universal quantifier	120 100%	26 **21.67%**	18 15.00%	4 3.33%	72 60.00%

Table 42.%: ognunu,a [human]

'I invite everybody…'		total	*good*	*bad*	*?*	*ø*
Iab.39a	*Chjamu à ognunu ...*	45 100%	29 **64.44%**	3 6.67%	2 4.44%	11 24.44%
Iab.39b	*Chjamu ø ognunu ...*	45 100%	20 **44.44%**	5 11.11%	1 2.22%	19 **42.22%**

Table 43.%: tutti [human]

	total	*good*	*bad*	*?*	*ø*
à tutti	75 100%	70 **93.33%**	1 1.33%	1 1.33%	3 4.00%
ø tutti	75 100%	6 8.00%	13 17.33%	3 4.00%	53 **70.67%**

Table 44.%: Indefinites [human]

	total	*good*	*bad*	?	ø
à + indefinite	314 100%	232 **73.89%**	13 4.14%	14 4.46%	55 17.52%
ø + indefinite	312 100%	73 **23.40%**	43 13.78%	20 6.41%	176 56.41%

Table 45.%: Indefinites [human]: *qualchissia, qualchidunu, qualcunu*

	total	*good*	*bad*	?	ø
à + indefinite	224 100%	174 **77.68%**	6 2.68%	9 4.02%	35 15.63%
ø + indefinite	223 100%	49 **21.97%**	30 13.45%	16 7.17%	128 **57.40%**

Table 46.%: unu [human]

	total	*good*	*bad*	?	ø
à unu	90 100%	58 **64.44%**	7 7.78%	5 5.56%	20 22.22%
ø *unu*	89 100%	24 26.97%	13 14.61%	4 4.49%	48 **53.93%**
alternative *ne* partitive + ø *unu/*	46 100%	5 10.87%			

Table 47.%: Verb mood: subjunctive *versus* indicative

'… who knows to repair my car?'		total	*good*	*bad*	?	ø
Iab.5a	*Cunnosci à unu chì sappii*[SUBJ] *riparà...*	45 100%	31 **68.89%**	3 6.67%	3 6.67%	8 **17.78%**
Iab.5b	*Cunnosci ø unu chì sappii*[SUBJ] *riparà...*	44 100%	11 **25.00%**	6 13.64%	2 4.55%	25 **56.82%**
Iab.5d	*Cunnosci à unu chì sà*[IND] *riparà...*	45 100%	27 **60.00%**	4 8.89%	2 4.44%	12 **26.67%**
Iab.5e	*Cunnosci ø unu chì sà*[IND] *riparà...*	45 100%	13 **28.89%**	7 15.56%	2 4.44%	23 **51.11%**

Table 48.%: unu [animate] (dog)

'Petru has struck one…		total	*good*	*bad*	?	ø
Iab.40a	*Petru hà tumbatu à unu …*	45 100%	13 **28.89%**	6 13.33%	1 2.22%	25 **55.56%**
Iab.40b	*Petru hà tumbatu ø unu …*	45 100%	18 **40.00%**	3 6.67%	2 4.44%	22 **48.89%**
Iab.40c/ alternative	*Petru ne hà tumbatu ø unu ...*	45 100%	11 **24.44%**			

Table 49.%: Negative quantifiers [human]

	total	*good*	*bad*	?	ø
à + negative quantifier	91 100%	78 **85.71%**	4 4.40%	1 1.10%	8 8.79%
ø + negative quantifier	91 100%	17 18.68%	13 14.29%	3 3.30%	58 **63.74%**

Table 50.%: nimu versus *nisunu*

'I see nobody'		total	*good*	*bad*	?	ø
Iab.21a	*Ùn vecu à nimu.*	46 100%	43 **93.48%**	2 4.35%	0 0%	1 2.17%
Iab.21b	*Ùn vecu ø nimu.*	46 100%	6 13.04%	8 17.39%	1 2.17%	31 **67.39%**
Iab.22a	*Ùn vecu à nisunu.*	45 100%	35 **77.78%**	2 4.44%	1 2.22%	7 15.56%
Iab.22b	*Ùn vecu ø nisunu.*	45 100%	11 **24.44%**	5 11.11%	2 4.44%	27 **60.00%**

Table 51.%: nulla [inanimate]

	total	*good*	*bad*	?	ø
à nulla	46 100%	4 **8.70%**	14 30.43%	1 2.17%	27 58.70%
ø nulla	46 100%	41 **89.13%**	0 0	0 0	5 10.87%

10   Bibliography

Abney, Steve. 1987. The English noun phrase in its sentential aspect. Cambridge, MA: Massachusetts Institute of Technology.

Ager, Dennis. 1999. *Identity, insecurity and image. France and language.* (Multilingual Matters 112). Clevedon: Multilingual Matters.

Aissen, Judith. 2003. Differential Object Marking: Iconicity vs. Economy. *Natural Language & Linguistic Theory* 21(3). 435–483. doi:10.1023/A:1024109008573.

Albertini, Jean. 1972. *Precis de grammaire corse*. Corti: CERC.

Alexiadou, Artemis, Liliane M. V. Haegeman & Melita Stauru. 2007. *Noun phrase in the generative perspective.* (Studies in Generative Grammar : SGG. - Berlin : De Gruyter Mouton, 1978- 71). Berlin: De Gruyter Mouton.

Arrighi, Jean-Marie. 1993. L'enseignement du corse depuis 1982. *Actes des Deuxièmes Journées Universitaires Corses de Nice 2*, 101–105. Nice: Centre d'Études corses de Nice.

Arrighi, Jean-Marie. 2002. *Histoire de la langue corse*. Paris: Gisserot.

Ascoli, Grazidio Isaia. 1882. L'Italia dialettale. *Archivio Glottologico Italiano* 8. 98–128.

Baker, Mark. 2015. *Case. Its Principles and its Parameters.* Cambridge: Cambridge University Press.

Benveniste, Émile. 1956. La nature des pronoms. In Morris Halle (ed.), *For Roman Jacobson*, 340–347. The Hague: Mouton.

Bertoni, Giulio. 1916. *Italia dialettale.* (Manuali Hoepli). Milano: Hoepli.

Blackwood, Robert J. 2004a. Corsican distanciation strategies: Language purification or misguided attempts to reverse the gallicisation process? *Multilingua - Journal of Cross-Cultural and Interlanguage Communication* 23(3). 233–255.

Blackwood, Robert J. 2004b. The Gallicisation of Corsica: The Imposition of the French Language from 1768 TO 1945. *Language Policy* 3(2). 133–152. doi:10.1023/B:LPOL.0000036179.70673.c1.

Blackwood, Robert J. 2007. L'Exception française? Post-war Language Policy on Corsica. *Journal of Multilingual and Multicultural Development* 28(1). 18–33.

Blackwood, Robert J. 2008. *The State, the activists and the islanders: language policy on Corsica.* (Language Policy : LAPO. - New York, NY : Springer, 2002- 8). Berlin: Springer.

Blackwood, Robert J. 2014. The top-down revitalization of Corsican: considering the reversal of a language shift in the linguistic and semiotic landscapes of Ajaccio. *French Studies* 68(1). 61–77.

Boberg, Charles. 2013. Surveys: the Use of Written Questionnaires in Sociolinguistics. In Christine Mallinson, Becky Childs & Van Herk (eds.), *Data Collection in Sociolinguistics. Methods and Applications*, 131–141. New York & London: Routledge.

Bossong, Georg. 1985. *Empirische Universalienforschung: differentielle Objektmarkierung in den neuiranischen Sprachen. .* Vol. 14. (Ars linguistica). Tübingen: Gunter Narr.

Bossong, Georg. 1991. Differential Object Marking in Romance and Beyond. In Dieter Wanner & Kibbee (eds.), *New Analyses in Romance Linguistics*, 143–170. (Current Issues in Linguistic Theory 69). Amsterdam & Philadelphia: Benjamins.

Bossong, Georg. 1998. Le marquage différentiel de l'objet dans les langues d'Europe. In Jack Feuillet (ed.), *Actance et valence dans le langues de l'Europe*, vol. 2, 193–258. (Empirical Approaches to Language Typology 20). Berlin: Mouton de Gruyter.

Bossong, Georg. 2008. *Die romanischen Sprachen: eine vergleichende Einführung.* (Studienbuch). Hamburg: Buske.

Bottiglioni, Gino. 1926. La penetrazione toscana e le regioni di Pomonte nei parlari di Corsica. *L'Italia dialettale* 2;3. 156–210; 1–69.

Bottiglioni, Gino. 1933-1942. *Atlante linguistico etnografico italiano della Corsica (ALEIC)*. Vol.

I–X. (L'Italia Dialettale). Pisa.

Bottiglioni, Gino. 1957a [1936]. La romanizzaazione nell'unità linguistica Sardo-Còrsa. *Sardegna Romana*, vol. 1, 126–139. Roma: Istituto di Studi Romani.

Bottiglioni, Gino. 1957b. *Miscellanea glottologica di Gino Bottiglioni, pubblicata in occasione del suo settantesimo compleanno da colleghi, amici, discepoli.* Modena: Società tipografica editrice modenese.

Bottiglioni, Gino. 1957c [1932]. Particolarità sintattiche dei dialetti còrsi (Dai materiali dell'Atlante Linguistico Etnografico Italiano della Corsica). *Corsica antica e moderna* 1. 120–124.

Boula de Mareüil, Philippe, Albert Rillard, Paolo Mairano & Jean-Pierre Lai. 2012. Corsican French questions: is there a prosodic transfer from Corsican to French and how to highlight it? *6th International Conference on Speech Prosody*, 418–421. Shanghai.

Brugè, Laura & Gerhard Brugger. 1994. On the accusative a in Spanish. 4(1). (University of Venice Working Papers in Linguistics). 3–44.

Carlson, Gregory Norman. 1977. Reference to Kinds in English. Amherst, MA: University of Massachusetts Amherst.

Casta, Santu. 2003. *La syntaxe du corse*. Ajaccio: CRDP Corse.

Chierchia, Gennaro. 1998. Reference to Kinfs across Languages. *Natural Language Semantics* 6.339–405.

Chiorboli, Jean. 1978. Reflets de la langue corse dans un manuscrit du XVIIème siècle. *Etudes corses*(10). 155–176.

Chiorboli, Jean. 1987. Traits morphosyntaxiques du corse: principaux constituants du syntagme nominal. *Études Corse* 28. 77–109.

Chiorboli, Jean. 1988. L'individuation corse dans la Romania intertyrrenique. *Publications universitaires de linguistique et d'antropologie (PULA)* 1. 61–88.

Chiorboli, Jean. 1991a. Langue corse: codification et contact linguistique. In Giovan Battista Pellegrini (ed.), *Fra dialetto e lingua nazionale: realtà e prospettive*, 13–21. Padova.

Chiorboli, Jean. 1991b. *La langue des Corses. Notes linguistiques et glottopolitiques*. Bastia: JPC Infograffia.

Chiorboli, Jean. 1992. Corse et Sicile: concordances intertyrrheniennes. *Bolletino. Centro di studi filologici e linguistici siciliani* 17. 147–180.

Chiorboli, Jean. 1997. Corse—italien. *Kontaktlinguistik Ein internationales Handbuch zeitgenössischer Forschung Contact Linguistics. An International Handbook of Contemporary Research Linguistique de contact Manuel international des recherches contemporaines*. Berlin & Boston: De Gruyter. http://www.degruyter.com/view/books/9783110151541.2/9783110151541.2.11.1214/9783110151541.2.11.1214.xml (27 April, 2014).

Chomsky, Noam. 1988. *Lectures on government and binding: the Pisa lectures*. Foris.

Chomsky, Noam. 1995. *The Minimalist Programm*. Cambridge, MA: MIT Press.

Chomsky, Noam. 2000. Minimalist Inquiries: The Framework. In Roger Martin, David Michaels & Juan Uriageraka (eds.), *Step by Step: Essays on Minimalist Syntax in Honor of Howard Lasnik*, 89–. Cambridge, MA: MIT Press.

Chomsky, Noam. 2001. Derivation by Phase. In Michael Kenstowicz (ed.), *Ken Hale. A Life in Language*, 1–52. Cambridge, MA: MIT Press.

Cinque, Guglielmo. 1999. *Adverbs and Functional Hears. A Cross-linguistic Perspective.* New York: Oxford University Press.

Colombani, Ignaziu. 1968. *Francorsu*. Bastia.

Comiti, Jean-Marie. 1992. *Les Corses face à leur langue. De la naissance de l'idiome à la reconnaissance de la langue*. Ajaccio: Squadra di u Finusellu.

Comrie, Bernard. 1989. *Language Universals and Linguistic Typology. Syntax and Morphology*. 2nd ed. Chicago: University of Chicago Press.

Contini, Michel. 1987. *Ètudes de géographie phonétique et de phonétique instrumentale du sarde.* . 2 vols. Alessandria: dell'Orso.

Corda, Francesco. 1983. *Saggio di grammatica gallurese*. Cagliari: Edizioni "3T."

Cowart, Wayne. 1997. *Experimental syntax: applying objective methods to sentence judgments*. Thousand Oaks, CA: Sage Publications.

Croft, William. 1988. Agreement vs. Case marking and Direct Objects. In Michael Barlow & Charles A. Ferguson (eds.), *Agreement in natural language. Approaches, theories, descriptions*, 159–179. Stanford: Center for the Study of Language and Information.

Croft, William. 1993. Case marking and the semantics of mental verbs. In James Pustejovsky (ed.), *Semantics and the lexicon*, 55–72. Dordrecht: Kluwer.

Croft, William. 2003. *Typology and universals*. Cambridge: Cambridge University Press.

Dahl, Östen & Kari Fraurud. 1996. Animacy in grammar and discourse. In Thorstein Fretheim & Jeanette K. Gundel (eds.), *Reference and Referent Accessibility*, 47–64. (Pragmatics & Beyond. New Series 38). Amsterdam & Philadelphia: Benjamins.

Dalbera-Stefanaggi, Marie-José. 1989. Le nouvel atlas linguistique de la corse et la banque de données linguistique corses: état de la recherche. *Études Corse* 32. 137–147.

Dalbera-Stefanaggi, Marie-José. 1991a. *Unité et diversité des parlers corses: le plan phonologique ; parenté génétique et affinité*. (Lingua, Cultura, Territorio. - Alessandria : Ed. Dell'Orso, 1981- 11). Alessandria: dell'Orso.

Dalbera-Stefanaggi, Marie-José. 1991b. Les Corses et leurs langues: sience et conscience. In Claude Martel (ed.), *Les Français et leurs langues: Colloque tenu à Montpellier les 5,6 et 7 settembre 1988*. Aix-en-Provence.

Dalbera-Stefanaggi, Marie-José. 1995. Collecte et traitements de données dialectales: Le Nouvel Arlas Linguistique de la Corse et la Banque de Données Langue Corse. *Géolinguistique. Bulletin du Centre di dialectologie* 6. 161–180.

Dalbera-Stefanaggi, Marie-José. 1997. Corsica. *The dialects of Italy*, 303–310. London & New York: Routledge.

Dalbera-Stefanaggi, Marie-José. 1999. Le corse-gallurien. *Géolinguistique. Bulletin du Centre di dialectologie* 8. 161–179.

Dalbera-Stefanaggi, Marie-José. 2001. Trois générations d'Atlas Linguistique de la Corse. *Essais de linguistique corse*, 27–35. Ajaccio: Alain Piazzola.

Dalbera-Stefanaggi, Marie-José. 2002. *La langue corse*. 1st ed. (Que Sais-Je?. - Paris : Presses Univ. de France, 1941- 3641). Paris: Presses Universitaires de France.

Dalbera-Stefanaggi, Marie-José. 2004. L'italianità linguistica della Corsica: pour une réévaluation des apports. *Etudes corses* 59. 55–63.

Dalbera-Stefanaggi, Marie-José. 2007a. From the linguistic atlas to the database, and vice versa: The Corsican example. In Yuji Kawaguchi, Toshihiro Takagaki, Nobuo Tomimori & Yoichiro Tsuruga (eds.), *Corpus-Based Perspectives in Linguistics*, 55–65. (Usage-Based Linguistic Informatics 6). Amsterdam & Philadelphia: Benjamins.

Dalbera-Stefanaggi, Marie-José. 2007b. *Nouvel atlas linguistique et ethnographique de la Corse: Aréologie phonétique*. nouv. éd. revue, et augm. Ajaccio & Paris: Piazzola & CTHS.

Damiani, Joseph. 1983. *Regards sur la langue corse*. (Lingua Corsa). Ajaccio: Terre Corse.

Damiani, Joseph. 1998. *Études linguistiques du corse*. Ajaccio: La Marge.

Danon, Gabi. 2006. Caseless nominals and the projection of DP. *Natural Language & Linguistic Theory* 24(4). 977–1008. doi:10.1007/s11049-006-9005-6.

Déchaine, Rose-Marie & Martina Wiltschko. 2002. Decomposing Pronouns. *Linguistic Inquiry* 33(3). 409–442.

Detges, Ulrich. 2005. La gramaticalización de los acusativos preposicionales en las lenguas iberorrománicas: Una hipótesis pragmática. In Gabriele Knauer (ed.), *Variación sintáctca en español. Un reto para las teorías de la sintaxis*, 155–173. Tübingen: Niemeyer.

Diesing, Molly. 1992. *Indefinites*. Cambridge, MA: MIT Press.

Diez, Friedrich Christian. 1836. *Grammatik der Romanischen Sprachen*. . Vol. 1. 4 vols. Bonn: Eduard Weber.

Dixon, Robert B. W. 1979. Ergativity. *Language* 55. 59–138.

Durand, Olivier. 2003. *La lingua còrsa: una lotta per la lingua*. (Studi Grammaticali e Linguistici. - Brescia : Paideia, 1950- 19). Brescia: Paideia.

Ettori, Fernand J. H. 1975. L'enseignement de la langue corse. *Langue française* 25. 104–111.

Ettori, Fernand J. H. 1979. A la recherche de la culture corse. Fille de l'Italie ou de la France? *Le mémorial des Corses*, 368–415. (Le Mémorial Des Corses 4). Ajaccio: Albiana.

Fabellini, Simona. 2010. *Sprachkonkurrenz auf Korsika vom 19. zum 20. Jahrhundert*. (Bonner Romanistische Arbeiten. - Frankfurt, a.M.: Lang-Ed, 1977- 101). Frankfurt a.M.: Peter Lang.

Fábregas, Antonio. 2013. Differential Object Marking in Spanisch: state of the art. *Borealis: An International Journal of Hispanic Lingustics* 2(2). 1–80.

Falcucci, Francesco Domenico. 1875. Saggi moderni: Corsica. In Giovanni Papanti (ed.), *I parlari italiani in Certaldo alla festa del V centenario di Messer Giovanni Boccacci*, 571–603. Livorno: Vigo.

Falcucci, Francesco Domenico. 1915. *Vocabolario dei dialetti, geografia e costumi della Corsica*. (Ed.) Pier Enea Guarnerio. . Vol. 1. (Biblioteca Della Società Storica Sarda). Cagliari: Società storica sarda.

Featherston, Samuel. 2007. Data in generative grammar: The stick and the carrot. *Theoretical Linguistics* 33(3). 269–318.

Filippi, Paul. 1992. Le français régional de Corse. Corti: Université de Corse.

Fusina, Jacques. 1994. *L'enseignement du corse. Histoire, développements récents et perspectives*. Aiacciu: A Squadra di u Finusellu.

Fusina, Jacques. 1999. *Parlons corse*. Paris: L'Harmattan.

Fusina, Jacques. 2000. *Corsican. The Corsican language in education in France*. 1st ed. (Regional Dossiers Series). Ljouwert & Leeuwarden: Mercator-Education.

Fusina, Jacques & Jean-Marie Arrighi. 2012. *Corsican. The Corsican Language in education in France (Mercator Regional Dossiers)*. 2nd ed. (Regional Dossiers Series). http://www.mercator-research.eu/research-projects/regional-dossiers/.

Gaggioli, Ghjaseppiu. 2012. *La langue corse en 23 lettres*. Aiacciu: Albiana.

García García, Marco. 2007. Differential object marking with inanimate objects. In Georg A. Kaiser & Manuel Leonetti (eds.), *Proceedings of the Workshop "Definiteness, Specificity and Animacy in Ibero-Romance Languages,"* vol. 122, 63–84. Konstanz: University of Konstanz.

Geach, Peter. 1962. *Reference and Generality.* Ithaca (NY): Cornell University Press.

Giacomo-Marcellesi, Mathée. 1978a. La langue: Différentiation micro-régionales et intercompréhension dans l'espace linguistique corse [chapitre IX]. *Pieve e Paesi: Communautés rurales corses*, 208–244. Paris: CNRS Édition.

Giacomo-Marcellesi, Mathée. 1978b. Etat des recherches en Linguistique Corse: II: La géographie linguistique. Le Nouvel Atlas Linguistique de la Corse. *Études Corse* 6(10). 139–154.

Giacomo-Marcellesi, Mathée. 1997. *Corse*. (Languages of the World 65). München & Newcastle: Lincom.

Giancarli, Pierre-Don. 2014. L'accusatif prépositionnel en corse. *Faits de Langues* 43. 197–212.

Gilliéron, Jules & Edmond Edmont (eds.). 1914. *Atlas linguistique de la France: Corse (ALF C)*. Vol. I–II. Paris: Champion.

Giorgi, Alessandra & Fabio Pianesi. 1997. *Tense and aspect: from semantics to morphosyntax*. (Oxford Studies in Comparative Syntax). New York, NY: Oxford University Press.

Giusti, Giuliana. 1993. *La sintassi dei determinanti*. (Rivista Di Grammatica Generativa. Serie

Monografica). Padova: Unipress.

Giusti, Giuliana. 1995a. Heads and Modifiers among Determiners: Evidence from Romanian. In Guglielmo Cinque & Giuliana Giusti (eds.), *Advances in Roumanian Linguistics*, 103–125. (Linguistik Aktuell 10). Amsterdam & Philadelphia: Benjamins.

Giusti, Giuliana. 1995b. A unified structural representation of (abstract) case and article. Evidence from Germanic. In Hubert Haider, Susan Olsen & Sten Vikner (eds.), *Studies in Comparative Germanic Syntax*, 77–93. (Studies in Natural Language and Linguistic Theory 31). Dordrecht: Kluwer Academic Publications.

Giusti, Giuliana. 1997. The categorial status of determiners. In Liliane Haegeman (ed.), *The new comparative syntax*, 95–123. (Longman Linguistics Library). London: Longman.

Giusti, Giuliana. 2009. On feature sharing and feature transfer. 19. (University of Venice Working Papers in Linguistics). 157–174.

Giusti, Giuliana. 2015. *Nominal Syntax at the Interfaces*. Venice: University Ca' Foscari.

Graziani, Paulu. 1929. Préface du tome 1 de l'Anthologie des écrivains corses. In Hyacinthe Yvia-Croce (ed.), *Anthologie des écrivains corses*, vol. 1, 7–14. Ajaccio: A Muvra.

Grimshaw, Jane. 1990. *Argument structure*. ([Linguistic Inquiry / Monographs] Linguistic Inquiry. - Cambridge, MA: MIT Press, 1976- 18). Cambridge, MA: MIT Press.

Guarnerio, Pier Enea. 1896. I dialetti odierni di Sassari, della Gallura e della Corsica (continuazione e fine). *Archivio Glottologico Italiano* 14. 385, 386.

Guarnerio, Pier Enea. 1904. *Il sardo e il còrso in una nuova classificazione delle lingue romanze*. (Archivio glottologico italiano 16). Torino: Ermanno Loescher.

Haspelmath, Martin. 2006. Terminology of case. In Andrei Malchukov & Andrew Spencer (eds.), the Oxford Handbook of Case. Oxford: Oxford University Press. https://www.eva.mpg.de/fileadmin/content_files/staff/haspelmt/pdf/CaseTerminology.pdf (24 October, 2020).

Heim, Irene. 1988. The semantics of definite and indefinite noun phrases. New York: Garland.

Helmbrecht, Johannes & Corinna Handschuh. 2016. Morphosyntactic coding of proper names and its implications for the Animacy Hierarchy. In Sonia Cristofaro & Fernando Zúñiga (eds.), Typological hierarchies in synchrony and diachrony. Amsterdam & Philadelphia: Benjamins.

Heusinger, Klaus von. 1997. Salience and definiteness. *Prague Bulletin of Mathematical Linguistics* 67. 5–23.

Heusinger, Klaus von. 2002. Reference and representation of pronouns. In Heike Wiese & Horst Simon (eds.), *Pronouns: Grammar and Representation*, 109–135. (Linguistik Aktuell 52). Amsterdam & Philadelphia: Benjamins.

Heusinger, Klaus von. 2004. Choice Functions and the Anaphoric Semantics of Definite NPs. *Research on Language and Computation*(2). 309–329.

Heusinger, Klaus von & Georg A. Kaiser. 2003. The Interaction of Animacy, Specificity, and Definiteness in Spanish. In Klaus von Heusinger & Georg A. Kaiser (eds.), *Processdings of the Workshop "Semantic and Syntactic Aspects of Specificity in Romanice Languages"*, 41–65. Konstanz: Universität Konstanz.

Heusinger, Klaus von & Georg A. Kaiser. 2005. The evolution of differential object marking in Spanish. *Proceedings of the Workshop "Specificity and the Evolution/Emergence of Nominal Determination in Romance,"* vol. 119, 33–69. Konstanz: University of Konstanz.

Heusinger, Klaus von & Georg A. Kaiser. 2007. Differential Object Marking and the lexical semantics of verbs in Spanish. In Georg A. Kaiser & Manuel Leonetti (eds.), *Proceedings of the Workshop "Definiteness, Specificity and Animacy in Ibero-Romance Languages,"* 85–110. Konstanz: Universität Konstanz.

Heusinger, Klaus von & Sofiana Chiriacescu. 2011. Pe-marked definite NPs in Romanian and discourse prominence. *Proceedings of the V NEREUS International Workshop "Mismatches*

in Romance", Zurich, October 8th/9th 2010, vol. 125, 33–54.

Higginbotham, James. 1983. Logical Form, Binding, and Nominals. *Linguistic Inquiry* 14, 395–420.

Holtus, Günter, Michael Metzeltin & Christian Schmitt (eds.). 1988. *Italienisch, Korsisch, Sardisch*. Tübingen: Niemeyer.

Hoop, Helen de. 2009. On (in)animate noun phrases. In Anastasia Giannakidou & Monika Rathert (eds.), *Quantification, Definiteness, and Nominalization*, 234–250. (Oxford Studies in Theoretical Linguistics). Oxford: Oxford University Press.

Hoop, Helen de & Andrej L. Malchukov. 2006. On fluid differential case marking: a bidirectional OT approach. *Lingua* 117(9). 1636–1656.

Hoop, Helen de & Andrej L. Malchukov. 2008. Case-marking strategies. *Linguistic Inquiry* 39(4). 565–587.

Hopper, Paul J. & Sandra A. Thompson. 1980. Transitivity in Grammar and Discourse. *Language* 56(2). 251–299.

Iemmolo, Giorgio. 2011. Towards a typological study of Differential Object Marking and Differential Object Indexation. Pavia: Università degli studi di Pavia.

Jaeggli, Osvaldo. 1982. *Topics in romance syntax*. (Studies in Generative Grammar). Dordrecht: Foris.

Jaffe, Alexandra M. 2001. The question of obligation: Authority and authenticity in Corsican discourse about bilingual education. In Monica Heller & Marilyn Martin-Jones (eds.), *Voices of Authority: Education and Linguistic Difference*, 269–296. Westport: Ablex.

Janich, Nina & Albrecht Greule (eds.). 2002. *Sprachkulturen in Europa: ein internationales Handbuch*. Tübingen: Narr.

Jones, Michael Allan. 1995. The prepositional accusative in Sardinian: its distribution and syntactic repercussions. In Martin Maiden & Mair Parry (eds.), *The dialects of Italy*, 336–384. London & New York: Routledge.

Kailuweit, Rolf. 2014. Avoiding typological affinity: negative borrowing asa strategy of Corsican norm finding. In Juliane Besters-Dilger, Cynthia Dermarkar, Stefan Pfänder & Achim Rabus (eds.): *Congruence in Contact-Induced Language Change*. Berlin & Boston: De Gruyter, 368-389.

Kailuweit, Rolf. 2015. Korsisch als Kompensationssprache. In Marío Alba Niño & Rolf Kailuweit (eds.), *Medien für Minderheitensprachen. Mediensprachliche Überlegungen zur Entwicklung von Minderheitensprachen*, 85–106. (Freiburger Romanistische Arbeiten 9). Freiburg: Rombach.

Keller, Frank. 1998. Grammaticality Judgments and Linguistics. Methodology. Research Paper EUCCS-RP-1998-3. Edinburgh, ms. http://www.google.de/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=0CDYQFjAA&url=http%3A%2F%2Fciteseerx.ist.psu.edu%2Fviewdoc%2Fdownload%3Fdoi%3D10.1.1.52.2559%26rep%3Drep1%26type%3Dpdf&ei=sp9DU7ToFIGMtAakgoHQCQ&usg=AFQjCNGog9Vat2E50lrhGoJzICgix_Yw7Q&bvm=bv.64367178,d.Yms (7 May, 2014).

Kloss, Heinz. 1967. Abstand languages and Ausbau languages. *Antropological Linguistics* 9(7). 29–41.

Kloss, Heinz. 1976. Abstandsprachen und Ausbausprachen. In Joachim Göschel, Norbert Nail & Gaston van der Els (eds.), *Zur Theorie des Dialekts: Aufsätze aus 100 Jahren Forschung*, 301–322. (Zeitschrift Fur Dialektologie and Linguistik, Beihefte 16). Wiesbaden: Steiner.

Kloss, Heinz. 1978. *Die Entwicklung neuer germanischer Kultursprachen seit 1800*. 2nd ed. (Sprache Der Gegenwart: Schriften Des Instituts Für Deutsche Sprache in Mannheim 37). Düsseldorf: Schwann.

Koch, Peter & Wulf Oesterreicher. 1985. Sprache der Nähe – Sprache der Distanz. Mündlichkeit und Schriftlichkeit im Spannungsfeld von Sprachtheorie und Sprachgeschichte.

Romanistisches Jahrbuch 36. 15–43.

Labov, William. 1972. *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press.

Laca, Brenda. 1995. Sobre el uso del acusativo preposicional en español. In Carmen Pensado (ed.), *El complemento direct proposicional*, 61–91. Madrid: Visor Libros.

Laca, Brenda. 2006. El objeto directo. La marcación preposicional. In Concepción Company Company (ed.), *Sintaxis histórica de la lengua española. Primera parte: La frase verbal*, vol. 1, 479–574. México: Fondo de la Cultura Económica y Universidad Nacional Autónoma de México.

Lamontagne, Greg & Lisa Travis. 1987. The syntax of adjacency. In *Proceedings of the 6th West Coast Conference on Formal Linguistics*. Stanford, CA: CSLI Publications. 173–186

Lausberg, Heinrich. 1963. *Romanische Sprachwissenschaft. Einleitung und Vokalismus*. 2nd ed. (Sammlung Göschen 128). Berlin: De Gruyter.

Lazard, Gilbert. 1998. Définition des actants dans les langues européennes. In Jack Feuillet (ed.), *Actance et valence dans les langues de l'Europe*, vol. 2, 11–146. (Empirical Approaches to Language Typology 20). Berlin: Mouton de Gruyter.

Ledgeway, Adam. 2011. Grammatica diacronica del napoletano. Tübingen: Niemeyer.

Loebel, Elisabeth. 1994. KP/DP-syntax: Interaction of case-marking with referential and nominal features. *Theoretical Linguistics* 20, 38–70.

Loi Corvetto, Ines & Annalisa Nesi. 1993. *La Sardegna e la Corsica*. Torino: UTET Libreria.

Longobardi, Giuseppe. 1994. Reference and Proper Names: A Theory of N-Movement in Syntax and Logical Form. *Linguistic Inquiry* 25(4). 609–665.

López, Luis. 2012. *Indefinite objects : scrambling, choice functions, and differential marking*. (Linguistic Inquiry / Monographs 63). Cambridge, MA: MIT Press.

Loporcaro, Michele. 2009. *Profilo linguistico dei dialetti italiani*. 1st ed. (Manuali Laterza 275). Roma: Laterza.

Lyons, Christopher. 1999. *Definiteness*. (Cambridge Textbooks in Linguistics). Cambridge: Cambridge University Press.

Malchukov, Andrej L. 2005. Case Pattern Splits, Verb Types and Construction Competition. In Mengistu Amberber & Helen de Hoop (eds.), *Competition and Variation in Natural Languages. The Case for Case*, 73–118. Amsterdam et al.: Elsevier.

Malchukov, Andrej L. 2006. Transitivity parameters and transitivity alternations: constraining co-variation. In Leonid Kulikov, Andrej L. Malchukov & Peter de Swart (eds.), *Case, valency and transitivity*, 329–359. Amsterdam & Philadelphia: Benjamins.

Malchukov, Andrej L. 2008. Animacy and asymmetries in differential case marking. *Lingua* 118(2). 203–221.

Mameli, Francesco. 1998. *Il logudorese e il gallurese*. (Collana Di Letteratura). Villanova Monteleone: Soter.

Manzini, Maria Rita & Leonardo Maria Savoia. 2006. *I dialetti italiani e romanci : morfosintassi generativa*. Alessandria: dell'Orso.

Marcellesi, Jean-Baptiste. 1981. Bilinguisme, diglossie, hégémonie: problèmes et tâches. *Langages* 61. 5–11.

Marcellesi, Jean-Baptiste. 1983. Identité linquistique, exclamatives et subordonnée: un modèle syntaxique spécifique en corse. *Etudes corses* 10(20/21). 399–423.

Marcellesi, Jean-Baptiste. 1984. La définition des langues en domaine roman; les enseignements à tirer de la situation corse. *Actes du XVIIe Congrès de Linguistique et Philologie Romanes*, vol. 5 (Sociolinguistique), 307–314. Aix-en-Provence.

Marcellesi, Jean-Baptiste. 1986. Le 'complement d'objet direct' en corse. *Morphosyntaxe des langues romanes.*, vol. 4, 127–138. (Actes Du XVIIe Congrès International de Linquistique et Philologie Romanes (Aix-En-Provence, 29 Août – 3 Settembre 1983)). Aix-en-Provence: Université de Provence.

Marcellesi, Jean-Baptiste & Jacques Thiers. 1988. Korsisch: Soziolinguistik/Sociolingistique. *Italienisch, Korsisch, Sardisch*, vol. 4, 809–819. (Lexikon Der Romanistischen Linguistik (LRL)). Tübingen: Niemeyer.

Marchetti, Pascal. 1974. *Le corse sans peine*. Chennevières-sur-Marne: Assimil.

Marchetti, Pascal. 1989. *La corsophonie: un idiome à la mer*. Paris: Albatros.

Marchetti, Pascal. 2001. *L'usu córsu*. Bastia: Sammarcelli.

Marchetti, Pasquale & Dumenicantone M. Geronimi. 1971. *Intricciate è cambiarine. Manuel pratique d'orthographe corse*. Nogent sur Marne: Beaulieu.

Mardale, Alexandru. 2008. Microvariation within Differential Object Marking. *Revue Roumaine de Linguistique* LIII(4). 448–467.

Martí, Luisa. 2008. The semantics of plural indefinite noun phrases in Spanish and Portuguese. *Natural Language Semantics* 16. 1–37.

Mattei, Antoine. 1864. La grammaire et le vocabulaire de l'idiome de la Corse. *Avenir de la Corse*. 2–3.

Matushansky, Ora. 2006. Head Movement in Linguistic Theory. *Linguistic Inquiry* 37(1). 69–109.

Maxia, Mauro. 1999. *Studi storici sui dialetti della Sardegna settentrionale*. Sassari: Studium ADF.

Maxia, Mauro. 2006. *Lingua, limba, linga: indagine sull'uso dei codici linguistici in tre comuni della Sardegna settentrionale*. (Samanunga 4). Cagliari: Condaghes.

Mayerthaler, Willi. 1981. *Morphologische Natürlichkeit*. (Linguistische Forschungen. – Frankfurt a.M.: Athenäum Verl, 1972- 28). Wiesbaden: Akademische Verl.-Ges. Athenaion.

Melillo, Armistizio Matteo. 1977. *Corsica*. (Ed.) Manlio Cortelazzo. (Centro Di Studio per La Dialettologia Italiana 5). Pisa: Pacini.

Merlo, Clemente. 1925. Concordanze corso-italiane centromeridionali. *L'Italia dialettale* 1. 238–251.

Meyer-Lübke, Wilhelm. 1890. *Grammaire des langues romanes. Phonetique. .* Vol. 1. Paris & Leipzig: Welter.

Meyer-Lübke, Wilhelm. 1900. *Grammaire des langues romanes. Syntaxe*. (Trans.) Georges Doutrepont. (Grammaire Des Langues Romanes / Meyer-Luebke, Wilhelm. - Paris, Leipzig : Welter, 1890-1906 ; 3 in 2 P.). Paris & Leipzig: Welter.

Milroy, James & Lesley Milroy. 1985. *Authority in language: investigating language prescription and standardisation*. (Language, Education and Society). London: Routledge & Kegan Paul.

Moravcsik, Edith. 1978. On the case marking of objects. In Joseph H. Greenberg, Charles A. Ferguson & Edith Moravcsik (eds.), *Universals of human language. Syntax*, vol. 4, 249–290. Stanford: Stanford University Press.

Muljačić, Žarko. 1989. Corsica. In Günter Holtus, Michael Metzeltin & Max Pfister (eds.), *La dialettologia italiana oggi. Studi offerti a Manlio Cortelazzo*, 287–312. (Tübinger Beiträge Zur Linguistik : TBL. - Tübingen : Narr, 1969- 335). Tübingen: Narr.

Myers, James. 2009. The design and analysis of small.scale syntactic judgement experiments. *Lingua* 119. 425–444.

Næss, Åshild. 2004. What Markedness marks: the Markedness Problem with Direct Objects. *Lingua* 114. 1186–1212.

Næss, Åshild. 2007. *Prototypical transitivity*. (Typological Studies in Language 72). Amsterdam & Philadelphia: Benjamins.

Næss, Åshild. 2012 [2008]. Varieties of Dative. In Andrej L. Malchukov & Andrew Spencer (eds.), *The Oxford Handbook of Case*. DOI: 10.1093/oxfordhb/9780199206476.013.0039.

Nesi, Annalisa. 1992. La Corsica. In Francesco Bruni (ed.), *L'Italiano nelle Regioni. Lingua nazionale e identità regionali*, vol. 1, 918–937. Torino: UTET.

Nesi, Annalisa. 1994. La Corsica. In Francesco Bruni (ed.), *L'Italiano nelle Regioni. Testi e documenti*, vol. 2, 895–911. Torino: UTET.

Nesi, Annalisa. 1998. Usi dell'italiano in Corsica durante l'Ottocento. In Alfieri, Gabriella & Arnold

Cassola (eds.), *La "lingua d'Italia": usi pubblici e istituzionali.* 113-128. Roma: Bulzoni.

Nesi, Annalisa. 2002. La Corsica. In Manlio Cortelazzo, Carla Marcato, Nicola De Blasi & Gianrenzo P. Clivio (eds.), *I dialetti italiani. Storia, struttura, uso*, 959–974. (La Nostra Lingua). Torino: UTET.

Neuburger, Kathrin. 2008. *Differentielle Objektmarkierung im Korsischen*. Freie Universität Berlin: unpublished master-thesis.

Neuburger, Kathrin & Elisabeth Stark. 2014. Differential Object Marking in Corsican: Regularities and triggering factors. *Linguistics* 52(2). 365–389.

Niculescu, Alexandru. 1959. Sur l'objet direct prépositionnel dans les langues romanes. *Recueil d'études romanes publié à l'occasion du IXe Congrès International de Linguistique Romane, Bucarest*, 167–185. Bucarest: Editions de l'Académie roumaine.

Oesterreicher, Wulf. 1993. Verschriftung und Verschriftlichung im Kontext medialer und konzeptioneller Schriftlichkeit. In Ursula Schaefer (ed.), *Schriftlichkeit im frühen Mittelalter*, 267–292. (ScriptOralia 53). Tübingen: Narr.

Ormazabal, Javier & Juan Romero. 2013. Differential Object Marking, Case and Agreement. *Borealis: An International Journal of Hispanic Lingustics* 2(2). 1–19.

Ottavi, Pascal. 2011. Regards sur le concept de diglossie, à l'épreuve du terrain corse. *Revue de linguistique et de didactique des langues (Lidil)*(44). 111–124.

Ottaviani, Thierry. 2010. *La Corse pour les nuls*. (Régions du France pour les Nuls). Paris: First-Gründ.

Petkanov, Ivan A. 1941. Appunti sui dialetti corsi e sardo-settentrionali. *Archivium Romanicum* 25. 192–200.

Piazza, François. 2009. Littératures. In Francis Pomponi, Geneviève Moracchini-Mazel, Jean-Marc Olivesi, Michel-Edouard Nigaglioni, Pierre Claude Giansily, Georges Ravis-Giordani, Marie-José Dalbera-Stefanaggi, François Piazza & Joseph Martinetti (eds.), *Corse*. Paris: Christine Bonneton.

Pollock, Jean-Yves. 1989. Verb Movement, Universal Grammar, and the Structure of IP. *Linguistic Inquiry* 20, 365–424.

Pomponi, Francis, Geneviève Moracchini-Mazel, Michel-Edouard Nigaglioni, Jean-Marc Olivesi & Pierre Claude Giansily. 2009. Histoire et art. *Corse*, 9–75. Paris: Christine Bonneton.

Popper, Karl R. 1959. *The Logic of scientific discovery*. London: Hutchinson.

RAE [Real Academia Española]. 2009. *Nueva gramática de la lengua española.* www.rae.es (last access 30.07.2020)

Rizzi, Luigi. 1997. The fine structure of the left periphery. In Liliane Haegeman (ed.), *Elements of Grammar*, 281–337. Dordrecht: Kluwer Academic Publications.

Rodríguez Mondoñedo, Miguel. 2007. The syntax of objects: agree and Differential Object Marking. Storrs: University of Connecticut.

Rohlfs, Gerhard. 1941. *L'Italianità linguistica della Corsica*. Wien: Schroll.

Rohlfs, Gerhard. 1949b. *Historische Grammatik der italienischen Sprache und ihrer Mundarten. Formenlehre und Syntax*. Vol. 2. 3 vols. (Bibliotheca Romanica 6). Bern: Francke.

Rohlfs, Gerhard. 1949a. *Historische Grammatik der italienischen Sprache und ihrer Mundarten. Syntax und Wortbildung*. Vol. 3. 3 vols. (Bibliotheca Romanica 7). Bern: Francke.

Rohlfs, Gerhard. 1984. *Von Rom zur Romania : Aspekte und Probleme romanischer Sprachgeschichte*. (Tübinger Beiträge Zur Linguistik). Tübingen: Narr.

Romani, Gilbert. 2005. *Grammaire corse pour le collège et l'école*. DCL Distribution Corse Du Livre (en ligne).

Rosch, Eleanor. 1978. Principles of categorization. In Eleanor Rosch & Barbara B. Lloyd (eds.), *Cognition and categorization*, 27–48. Hillsdale: Erlbaum.

Salvioni, Carlo. 1916. Note di dialettologia corsa. *Rendiconti dell'Istituto Lombardo di Scienze e Lettere* 49. 705–880.

Schikowski, Robert & Giorgio Iemmolo. 2015. Commonalities and differences between differential object marking and indexing. Universität Zürich, ms. https://www.google.de/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&cad=rja&uact=8&ved=0ahUKEwjpjMz77YHWAhWMalAKHbChDxoQFggyMAA&url=http%3A%2F%2Fwww.ivs.uzh.ch%2Fschikowski%2Fpublications%2F2015-dom-vs-doi.pdf&usg=AFQjCNFbFQKmUz7bda9iMfywv-Cy3HItlw (31 August, 2017).

Schütze, Carson T. 1996. *The empirical base of linguistics: grammaticality judgments and linguistic methodology*. Chicago: University of Chicago Press.

Schütze, Carson T. & Jon Sprouse. 2014. Judgment data. In Robert J. Podesva & Devyani Sharma (eds.), *Research Methods in Linguistics*, 27–50. Cambridge: Cambridge University Press.

Siewierska, Anna. 1997. The formal realization of case and agreement marking: a functional perspective. In Anne-Marie Simon-Vandenbergen, Kristin Davidse & Dirk Noël (eds.), *Reconnecting Language*, 181–210. Amsterdam: Benjamins.

Silverstein, Michael. 1976. Hierarchy of Features and Ergativity. In Robert B. W. Dixon (ed.), *Grammatical Categories in Australian Languages*, 112–171. Canberry & New Jersey: Australian Institute of Aboriginal Studies & Humanities Press.

Sinnemäki, Kaius. 2014. A typological perspective on Differential Object Marking. *Linguistics* 52(2). 281–313.

Spolsky, Bernard. 2004. *Language Policy*. (Key Topics in Sociolinguistics). Cambridge: Cambridge University Press.

Stark, Elisabeth. 2006. *Indefiniteheit und Textkohärenz*. Tübingen: Niemeyer.

Stowell, Tim. 1989. Subjects, Specifiers, and X-bar Theory. In Baltin, Mark R. & Anthony S. Kroch (eds.), *Alternative Convepctions of Phrase Structure*. Chicago: University of Chicago Press.

Swart, Peter de. 2006. Case markedness. *Case, Valency and Transitivity*, 249–267. (Studies in Language Companion Series 77). Amsterdam & Philadelphia: Benjamins.

Swart, Peter de. 2007. Cross-linguistic Variation in Object Marking. Nijmegen: University of Nijmegen.

Tagliavini, Carlo. 1972 [1949]. *Le origini delle lingue neolatine*. 6. ed., interamente rielaborata ed. aggiornata con 51 fig. e 98 ritratti. Bologna: Pàtron.

Thiers, Ghjacumu. 1987. L'enquête épilinguistique. *Études Corse* 28. 39–75.

Thiers, Jacques. 1986. Épilinguisme, élaboration linguistique et volonté populaire, trois supports de l'individuation sociolinguistique corse. *Langages*(83). 65–74.

Thiers, Jacques. 1993. Language contact and Corsican polynomia. In Rebecca Posner & John N. Green (eds.), *Billingualism and Linguistic Conflict in Romance*, vol. 5, 253–270. (Trends in Romance Linguistics and Philology 71). Berlin & New York: Mouton de Gruyter.

Thiers, Jacques. 1997. Français—corse. In Hans Goebl, Peter H. Nelde, Zdenek Stary & Wolfgang Wölck (eds.), *Kontaktlinguistik. Ein internationales Handbuch zeitgenössischer Forschung*, 1203–1213. Berlin & Boston: De Gruyter. http://www.degruyter.com/view/books/9783110151541.2/9783110151541.2.11.1201/9783110151541.2.11.1201.xml (27 April, 2014).

Thiers, Jacques. 2008 [1989]. *Papiers d'identité(s)*. 2nd ed. Aiacciu: Albiana.

Torrego, Esther. 1998. *The dependencies of objects*. ([Linguistic Inquiry / Monographs] Linguistic Inquiry. - Cambridge, Mass. : MIT Press, 1976- 34). Cambridge, MA: MIT Press.

Travis, Lisa. 1984. *Parameters and Effects of Word Order Variation*. Dissertation. MIT.

Trudgill, Peter. 1974. *The social differentiation of English in Norwich*. (Cambridge Studies in Linguistics 13). Cambridge: Cambridge University Press.

Viale, Salvatore. 1817. *Dionomachia. Poemetto eroi-comico con note.* London.

Wandruszka, Ulrich. 2011. Präpositionaler Akkusativ und Satzakzent. *Zeitschrift für romanische Philologie* 127(2). 328–351. doi:10.1515/zrph.2011.022.

Watanabe, Akira. 2000. Feature copying and binding: evidence from complementizer agreement

and switch reference. *Syntax* 3(3). 159–181.

Wegener, Heide. 1991. Der Dativ – ein struktureller Kasus? In Gisbert Fanselow & Sascha W. Felix (eds.), *Struckturen und Merkmale syntaktischer Kategorien.* Tübingen: Nass. 70–103.

Weinreich, Uriel. 1954. Is a structural dialectology possible? *Word* 10. 388–400.

Wiltschko, Martina. 2014. The universal structure of categories. Towards a formal typology. Cambridge: Cambridge University Press.

Yamamoto, Mutsumi. 1999. *Animacy and Reference*. (Studies in Language Companion Series 46). Amsterdam & Philadelphia: Benjamins.

Yvia-Croce, Hyacinthe. 1987 [1929]. *Anthologie des écrivains corses*. Ajaccio: Cyrnos et Méditerranée.

Zamparelli, Roberto. 2000. *Layers in the determiner phrase.* Garland: New York.

Deutsche Zusammenfassung/English Summary

1. Deutsche Zusammenfassung

Die Dissertation **Differential Object Marking in Corsican. Distribution, triggers, functions** *(Differentielle Objektmarkierung im Korsischen. Distribution, Auslöser, Funktionen)* beschäftigt sich mit der Differentiellen Objektmarkierung im Korsischen.

Das Phänomen: Die Differentielle Objektmarkierung (DOM) beschreibt die overte morphologische Markierung – im Korsischen durch das pränominale Morphem *a* – bestimmter (nicht aller) direkter Objekte abhängig von Eigenschaften des Objektnomens und/oder seiner direkt umgebenden Struktur aufgrund von semantischen, syntaktischen oder pragmatischen Kriterien. Als Auslöser für DOM spielen meist Belebtheit und/oder Referentialität des Objekt-Nomens eine zentrale Rolle. Das trifft auf zahlreiche DOM-Systeme in den romanischen Sprachen zu und so auch auf das korsische.

Die Daten: Für den Fragebogen habe ich Sätze aus der Literatur zur korsischen DOM übernommen, modifiziert und ergänzt. Mit Hilfe des Fragebogens habe ich im Laufe zweier Feldforschungsreisen (im Juni 2010 und September 2011) 46 Sprecher*innen des Korsischen befragt. Die dafür gewählte Methode *forced choice* ('erzwungene Wahl') erfordert von den Informant*innen, aus den angebotenen Optionen die für sie 'beste Option' auszuwählen. Damit sollte die bis dato geringe und lückenhafte Datenlage auf systematische Weise ergänzt werden.

Die Ergebnisse: Die Verteilung der gegebenen Antworten fiel sehr heterogen aus. Kurz gesagt, orientiert sich die korsische DOM in der Wahrnehmung der Informant*innen nicht entlang feststehender Nominalklassen und reagiert auch nicht eindeutig auf die genannten Merkmale Belebtheit und Definitheit: keine der Klassen und keiner der Einzelsätze hat alle 'best option'-Antworten auf sich vereint. Dennoch sind die beiden Merkmale Menschlichkeit und Definitheit relevant, wenn auch, je nach nominaler Klasse, zu unterschiedlichen Teilen. Pronomina[human] werden immer *a*-markiert, Eigennamen sowohl als auch, lexikalische Nomen tendenziell nicht. Allgemein lässt sich sagen, dass Menschlichkeit des Objektreferenten und/oder ein hoher Grad an Definitheit eine morphologische Akkusativmarkierung triggert. In struktureller Hinsicht hat sich als zentrale Eigenschaft der korsischen DOM die Inkompatibilität des definiten Artikels und des DOM-Markers *a* herausgestellt.

    Dieser Inkompatibilität habe ich die Annahme zugrunde gelegt, dass der DOM-Marker und der definite Artikel im Zusammenhang mit referentiellen, definiten und/oder menschlichen direkten Objekten im Korsischen die gleiche Funktion übernimmt: Beide Elemente markieren Kasus. Ob Objekte nun mit dem definiten Artikel oder mit *a* markiert werden, hängt von deren Struktur ab: Pronomina können nie mit definiten Artikeln markiert werden, sie erhalten die *a*-Markierung. Eigennamen treten sowohl mit *a*-Markierung oder definitem Artikel auf und Nomen erhalten ausschließlich den definiten Artikel.

Zusätzlich enthält die Dissertationsschrift auch ein Kapitel zur Sprachgeschichte des Korsischen sowie seiner Klassifizierung als italoromanische Sprachvarietät. Außerdem bietet das Kapitel zum Stand der Forschung auch einen Einblick in die Wissenschaftsgeschichte der korsischen Sprachwissenschaft, deren Beginn als eigenständige Disziplin im deutsch-sprachigen Raum mit dem Eintrag ins Lexikon der Romanischen Linguistik (LRL) 1988 zusammenfällt.

Die Arbeit leistet somit einen Beitrag sowohl zur romanischen und korsischen Sprachwissenschaft als auch zur allgemeinen Forschung des Phänomens DOM.

262

2. English Summary

The dissertation ***Differential Object Marking in Corsican. Distribution, triggers, functions***, approaches the phenomenon of Differential Object Marking in Corsican.

The phenomenon: Differential Object Marking (DOM), in its most basic and traditional conception, is the marking of certain but not all nominals in direct object (DO) function according to different semantic, syntactic and pragmatic criteria by means of an overtly expressed morphosyntactic marker. In Corsican, DOM is expressed by the morpheme *à*, which is homophonous to the dative and locative prepositions *à*, and appears with the DO if certain semantic, i.e. humanness and/or syntactic conditions, no overt definite article, apply and in turn, does not surface if these do not apply.

The data: My approach is based on a corpus of empirical language data which I collected during two fieldtrips in June 2010 and September 2011. By means of a written questionnaire, I surveyed 46 informants, using a *forced choice* task.

The results: The answers of the informants have been quite heterogeneous. Corsican DOM is not categorial, neither in terms of noun classes, nor in terms of semantic features, such as humanness and/or definiteness. Nevertheless, these features have an impact on the marking pattern, but to divergent degrees: proforms[human] get consistently *a*-marked but common nouns[human] bear the definite article and no *a*-marker. Among the class of proper nouns not only human but also inanimate DOs get *a*-marked, e.g. toponyms. In general, a high degree of animacy (humanness) and/or definiteness favours overt marking.

In structural terms, a key property of Corsican DOM is the incompatibility of the definite article and the DOM-marker: e.g. common nouns with a definite article are not *a*-marked.

I assume that, in Corsican, definite and/or human DOs require overt case marking expressed either with the definite article or the *a*-marker: both items may function as overt accusative case markers. If the DOs get marked either by *a* or the definite article depends on the nominal: nominals which never appear with definite articles such as e.g. proforms, get *a*-marked, as well as (certain) proper names; common nouns never bear *a*-marking but most consistently the definite article.

In addition to Corsican DOM, I provided also information about the history of Corsican and its speakers and gave some insights in the history of the science regarding Corsican as an independent field of linguistics.

This dissertation contributes to Corsican and Romance linguistics as well as to the research on the phenomenon of DOM.

CV

Kathrin Anne Neuburger studied Italian Philology and History of Arts at the Ruprecht-Karls-Universität, Heidelberg, the Università Roma Trè and at the Universidad de la Habana. She graduated at Freie Universität Berlin (Magister Artium) in 2008.

In 2009 she started her PhD as a member of the Doctoral Program at the University of Zurich, in 2010 she got a grant from Evangelisches Studienwerk Villigst e.V. She has been to Cambridge, UK, as a visiting student and conducted fieldwork in Corsica. The PhD-project was presented on many different national and international conferences, such as e.g. *Sky Symposium Case in and across language,* Helsinki, *XXXI Romanistentag,* Bonn, *Conference on Differential Object Marking,* Tromsø, *Cambridge Italian Dialect Syntax-Morphology Meeting*, Cambridge (UK) and *XXXV Romanistentag,* Zürich.

Statement of Authorship/Eigenständigkeitserklärung

I certify that this dissertation is my own work and contains no material which has been accepted for the award of any degree or diploma in any academic institute. Moreover, to the best of my knowledge and belief, it contains no material previously published or written by another person, except where due reference is made in the text.

Ich erkläre hiermit, dass diese Dissertation nicht bereits im Rahmen eines anderen Promotionsverfahrens ganz oder in Teilen angenommen oder abgelehnt wurde. Außerdem erkläre ich, dass ich die vorliegende Dissertation selbständig und ohne fremde Hilfe verfasst, andere als die angegebenen Quellen nicht benützt und die den benützten Quellen wörtlich und inhaltlich entnommene Stellen als solche kenntlich gemacht habe.

Berlin, den 10. Oktober 2020   _____

                                                                        Unterschrift