

# Syntactic Protocols to Enhance Inclusive Cultural Identity. A Case Study on Istro-Romanian Clausal Structure\*

*Giuliana Giusti and Iulia Zegrean*

Università Ca' Foscari di Venezia

(<giusti@unive.it>; <iuliazegrean@gmail.com>)

## *Abstract:*

This paper has a general and a specific goal. The general goal is quite ambitious and consists in proposing a new linguistic approach, named Protocol Linguistics, that should encompass points of division among linguists of different theoretical persuasion and permit a common effort to put the most recent advances in linguistics at the service of general interests, such as language policies, language education, language rehabilitation, endangered language documentation, and many more third mission type of environments. The second specific goal is to provide an example of a possible protocol for the documentation of a severely endangered language, Istro-Romanian, a variety of eastern Romance spoken in the Istrian peninsula in Croatia, also named Vlački or Žejanski. The general aim is to suggest that language awareness based on knowledge of syntax, which highlights similarities as well as differences and conceives the differences as variation among a small range of choices, can ground the construction of inclusive cultural identity, which will enhance social cohesion as well as the preservation of minority languages.

*Keywords:* Endangered Languages, Istro-Romanian, Language Awareness, Metalinguistic Competence, Protocol Linguistics

\* This work is empirically based on Zegrean (2012). Iulia Zegrean did most of her research with a grant from the Ca' Foscari University of Venice, and during a semester at the Center for Advanced Studies in Theoretical Linguistics (CASTL), University of Tromsø. We thank the following consultants Livija (and Redento) Fable, her sisters Nelja and Vanda, Mauro and Robert Dorčić. Giuliana Giusti elaborated the protocols on a research grant awarded by CRASSH – University of Cambridge (26.09.2012-2.12.2012). We are also indebted to Zvezdana Vrzić and two anonymous reviewers for constructive criticism. For the sake of legal requirements, Iulia Zegrean is responsible for Sections 3.1-4, and Giuliana Giusti for the remaining parts of the paper.

## *1. Introduction*

### *1.1 Aims and Structure of the Paper*

The human language capacity interacts with almost every single aspect of human behavior. The study of language from the internal linguistic perspective is however conducted as a niche discipline whose recent advances are rarely disseminated into other fields. As a consequence, the considerable headway made by linguistic theories since the beginning of the last century has had little impact on social phenomena. Linguists are apparently skeptical as regards the sociological impact of their work. And they are also unable to join in a common effort, abandoning technicalities that are necessary at highly specialized levels of discussion but constitute an obvious impediment in establishing a dialogue with scholars working in other sciences and decision takers acting on political and social issues.

We believe that knowledge of the dual nature of language between the biological and cultural dimensions is crucial to take decisions and plan language education and teaching, rehabilitating techniques, compensatory tools for disabled subjects, documentation and preservation of minority and heritage languages, to mention just a few areas that have great impact on the society.

This paper aims to set the foundations of a metamodel of linguistic research, called Protocol Linguistics, that can be shared by linguists of different empirical specializations and theoretical persuasion and be accessible to the non-linguistic World. The proposed syntactic model is declined in “protocols”, which are descriptive properties of natural languages, organized in clusters. The paper will tackle the issue from one of the many possible areas in which linguistic knowledge could have strong impact; namely, documentation of endangered languages.

Minority groups are subdue to constant tension between cultural isolation and assimilation. Awareness of the multifaceted form of the language faculty which makes human beings naturally multilingual can help the construct of identities that are not based on mutually exclusive features, but on the recognition of the differences conceived as manifestations of broader, more abstract common properties. Documentation of a language in serious danger and the hope of preservation can be supported by disseminating linguistic metacompetence on language types and families, language contact, language history and change, bilingualism and acquisition. In this perspective, we propose a case study on an eastern Romance language, Istro-Romanian, also named Vlaški-Žejanski, located in Croatia not far from the Italian border.

The structure of the paper is as follows. The rest of this Section grounds the claim that linguistic metacompetence is necessary to construct a special type of language awareness that can build inclusive (as opposed to exclusive)

cultural identity. Section 2 sketches the notion of Protocol Linguistics, with special reference to syntax. Section 3 provides an example of the application of a few protocols to Istro-Romanian. Section 4 spells out the possible results of such protocols to different actions that can be taken, all converging to construct language awareness to enhance inclusive cultural identities.

### *1.2 Linguistic Metacompetence and Inclusive Cultural Identity*

Minority languages run two very dangerous tendencies that are experienced at every corner of the Globe. On the one hand, lack of prestige leads to abandonment of the minority language in favor of the majority language, with or without consequent loss of cultural identity, but with secure loss of one of its most important components; that is the language. This is also due to general biases against bilingualism and a misconceived notion of native competence. On the other hand, the pride set on the cultural values of local varieties can lead to refusal of the prestige or standard language with the result of creating linguistic and cultural ghettos, which marginalize their components and doom the minority language to poverty of registers, limited to informal usages and slang, and ultimately also leading the language to extinction.

As usual, knowledge is the only means for cultural enhancement. Knowledge of the general properties of language and of the specific properties of the numerous languages each speaker masters can, in our opinion, be a solid pillar for the construction of cultural identity and language documentation and preservation. The language awareness we want to raise is based on linguistic metacompetence, namely on knowledge of language, conceived as a complex human cognitive capacity, a biological endowment interacting with the social nature of human beings.

The linguistic research on the impact of language in the formation of cultural identities (cf. Edwards 2009, for an introduction) has up to now only dealt with the study of accents and vocabulary, almost completely disregarding syntactic matters, which are instead crucial to understand language history, contact and attrition, as well as L1 and L2 acquisition.

Formal syntactic research, and in particular the theory known as generative grammar, aims at capturing the general properties of language that are inborn and common to all human beings (principles), and the restricted possibility of variation that is also inborn and needs to be set for one choice or another (parameters), through exposure to linguistic input, which is necessarily varied and mixed. How parameters are set and what consequences the exposure to a mixed input can have are two issues, particularly relevant to the study of languages situated in transregional and transcultural environments. Settling down these issues in a form, accessible to the general public and open to multidisciplinary enrichment can in our opinion be the ground

for a particular kind of language awareness enhancing the construction of inclusive cultural identity.

Awareness should be at the base of positive cultural identity. This is true in particular if awareness of specific aspects of one's culture is based on sound argumentation and not on biased pride or despise of competing cultures and languages. For example, the notion of linguistic ancestor as proof of cultural descend may become a sensitive one, especially if there are no clear historical witnesses in one sense or another.

A related issue is the fact that multiculturalism and in particular multilingualism was and still is the norm. The very idea of a "pure" cultural identity formed across time and holding in the present is a myth that finds no justification in either linguistic or psycho-sociological perspective. Only very secluded tribes develop isolated cultures, which are usually conservative and underdeveloped. Contacts and contaminations are the basis of economical, cultural and technological development. Language change goes parallel to cultural change in that the human being is naturally multilingual and the languages/varieties coexisting in a complex society naturally influence each other in interesting ways.

The kind of awareness we are aiming at is awareness of a cognitive capacity (the human language faculty) and how one's language(s) can be situated in a vast but not infinite frame of linguistic variation. It can be declined along three different dimensions: (i) awareness of language as a human capacity manifested in languages, varieties, dialects, which are multi-systems formed by and including coexisting subvarieties, differentiated at different degrees of inclusiveness; (ii) awareness of the multilingual nature of the individual, understood as the cognitive capacity of unconscious competence of more than one variety, dialect, sociolect, or standard language, which makes a single person a member of multi-layered groups at different degrees of inclusiveness; and finally, (iii) awareness of differences and similarities across the different coexisting languages / varieties in the competence of the same individual. Awareness of these three aspects of language can, in our opinion, enhance the construction of language identity in an inclusive fashion, and help eliminate or at least constrain the many ethnic and social biases conveyed by language. This kind of awareness, we propose, can be achieved through knowledge of basic properties of language in general and of particular languages. Such properties can be captured in the forms of protocols, which we describe in the following Section.

## *2. Protocol Linguistics*

A protocol is an established procedure which applies in the same way with the same tools in different but comparable situations. It therefore permits to avoid interference that may cause problems of different types while pursuing

an objective. When the objective regards the acquisition of information, it ensures that such information is comparable. The language awareness that we aim to build is grounded on comparison. A linguistic feature is described as a general principle declined into a restricted number of variation possibilities. Comparison between two or more languages can highlight differences and similarities between those languages. If a linguistic difference is conceived as a restricted variation of choice inside one and the same open variable, the cultural identity constructed on language awareness will have an inclusive nature. In other words, linguistic diversity can be conceived as unifying and not as dividing.

It would be unfair to deny that the present proposal stems from the principle and parameters framework, but also typological linguistics has implication patterns that suggest a similar unifying perspective. We conceive a protocol feature not just as a single proposition, to be valued as true or false (present or absent) in a language, but as a cluster of related properties. This is different from the approach taken in other inventories of language features such as WALS<sup>1</sup> or SSWL.<sup>2</sup> Going from the more general to the more specific, we can construct protocols with growing clustering complexity and language specificity, as will be detailed in this Section.

Establishing the relevant features for the relevant languages produces a simple table-chart with the languages that are relevant to our search on the horizontal axis and the properties that are going to be tested on the vertical axis and valued dichotomously with a +/- value.

### *2.1 The Horizontal Axis: the Relevant Languages*

As anticipated above, we present a case for the proposed protocol methodology applying it to a severely endangered eastern Romance variety<sup>3</sup> spoken in the Croatian peninsula of Istria, called Istro-Romanian by the philological-linguistic tradition and Vlaški-Žejanski by the communities speaking the language.

The population migrated to the geographical area of Mount Učka before the 16<sup>th</sup> century (cf. Pușcariu 1926), but it is hardly clear to historians, linguists or anthropologists when exactly they arrived in the region nor where they originate from. Today, the speakers of Istro-Romanian seem to constitute the smallest (ethno-)linguistic community in Europe. The majority of them are elderly or middle-aged, very few children have even passive competence of it.

<sup>1</sup> <<http://wals.info/>> (09/2015).

<sup>2</sup> <<http://sswl.railsplayground.net/>> (09/2015).

<sup>3</sup> We deliberately use the terms “variety”, “dialect”, and “language” as synonyms, in the sense of “manifestations of the human faculty of language”.

This linguistic variety is classified as *seriously endangered* (UNESCO Atlas of World's Languages in Danger).<sup>4</sup> In 2007, it was included on the *List of protected intangible cultural heritages* by the Ministry of Culture of the Republic of Croatia.

The debate as to the origin and classification of Istro-Romanian is, for many (linguists and non-linguists), a central topic. There are two more acknowledged theories. One claims that Istro-Romanian is one of the four “historical dialects” of the Romanian language arising from Proto-Romanian, on a par with Daco-Romanian (present day Romanian), Aromanian and Megleno-Romanian. The other claims that it detached from the common linguistic trunk later than the other idioms, thus it is simply a Daco-Romanian dialect (Kovačec 2009).

Istro-Romanian is of particular interest to show how Protocol Linguistics can enhance inclusive cultural identity. There are at least three dimensions of linguistic research which relate Istro-Romanian to other better studied varieties, avoiding a fossilization of the issue on the Daco-Romanian vs. Proto-Romanian ancestor, which cannot be settled in view of lack of documentation.

The first is the inter-Balkan dimension. Croatian and Romanian are Balkan languages at different degrees (Romanian being one of the “most Balkan”, Croatian being one of the “less Balkan” ones, cf. Mišeska-Tomić 2006 for an overview). The issues that can be settled are the following: What are the Balkan features present in Croatian and Romanian? Are these features also present in Istro-Romanian? For those features that are present/absent, do they come from (a previous stage of) Proto-Romanian and were reinforced/weakened by the contact with Croatian? Are there Balkan features in Istro-Romanian that are not present in either Romanian or Croatian but are present (for example, in Aromanian and Megleno-Romanian, which are considered by Joseph 1999 as “more Balkan” than Romanian)? Zegrean (2012) gives an overview of the Balkanisms present in Istro-Romanian with respect to the other three Balkan Romance languages. Her conclusion is that Istro-Romanian is the “least Balkan” of all. This is not surprising, considering the geographical position and the contact with Croatian.

A second obvious line of interest raised by the study of Istro-Romanian is the intra-Romance dimension: What Romance features of Romanian are present/absent in Istro-Romanian? In the perspective of a partial contact with the Italo-Romance varieties present in the Istrian peninsula until the last century, would it be correct to hypothesize any influence or contact that these varieties may have induced on Istro-Romanian in the preservation of certain Romance properties absent in Balkan-Romance?

<sup>4</sup> <<http://www.unesco.org/languages-atlas/>> (09/2015).

A third perspective is provided by the alloglottal dimension: Croatian is only marginally Balkan; on the contrary, it has many Slavic features that are not present in Balkan languages. What are the Slavic features of Croatian also present in Istro-Romanian? The alloglottal perspective can also be considered from a broader point of view and raises two other questions: Are there Slavic features in Istro-Romanian that are not present in the neighboring varieties (Croatian, Chakavian)? Are there Romance features in Istro-Romanian that are not present in Italian or in Romanian?

In this paper we provide no direct answer to any of these questions. The protocols are designed to raise empirical questions that can inspire elicitation tests, interviewing techniques, and a system of corpus annotation that will in the long run give us more precise results.

The horizontal axis of our tentative chart for Istro-Romanian will therefore include Romanian, Croatian, and Italian (including their non-standard varieties).

## 2.2 *The Vertical Axis: the Clusters of Feature*

The vertical axis is much more dependent on how deep our survey is intended to be. For expository purposes, we can only be quite sketchy at this stage. Our prototype protocol starts from the core properties of sentence structure.

The subject-predicate dichotomy is acknowledged by all linguistic theories. Another uncontroversial property is that in many languages the verbal element expressing finite Mood (and Tense) also agrees for the person features of the subject. Furthermore, rich verbal inflection for the person of the subject often correlates with the possibility of omission of pronominal subjects in finite clauses. Finally, the Subject Verb order can vary across languages (cf. Dryer 2013). The first cluster of properties for the comparison of the four languages at stake includes: Subject-Verb Agreement, SV/VS alternation, null Subject pronouns in finite clauses.

Auxiliary insertion to express past Tense or perfective Aspect is generally found across European languages. The two possible auxiliaries *be* and *have* select a past participle form of the verb, which agrees with the subject for the gender and number in the presence of auxiliary *be* only. The second cluster of properties we suggest can be formulated as: Auxiliary + Past participle to express past Tense/perfective Aspect; Auxiliary alternation according to V-classes; Auxiliary *have*-past participle agreement; Auxiliary *be* + past participle agreement.

The position of the verb in the clause is the pivot of a number of properties such as its position with respect to adverbs, the possibility of V-fronting to the left of an auxiliary (which is present in many Balkan languages, and not present in Romance), and the possibility of VP-deletion in short answers (which is absent in Romance but present in Slavic).

Great variation across the languages of Europe is found as regards the position of clitics in the clause, which can appear to cluster around the Tense position, or to occupy the so-called Wackernagel position (the second position in the clause). The clitic nature is in general reserved to functional elements such as pronouns, auxiliaries, or negative particles. A tentative protocol for this intricate matter must therefore regard these two dimensions.

In the following Section we give a first formulation of the partial protocols for Istro-Romanian providing some tentative answers mainly based on fieldwork conducted by Iulia Zegrean in 2009-2010.

### *3. A Partial Protocol for Istro-Romanian Clausal Structure*

Before starting the presentation of the data, a caveat on orthography is at stake. We do not take a position on how to spell Istro-Romanian. Vrzić (2009) makes a recent proposal grounded on the spelling system of Croatian, motivated by the fact that the people who speak it are not used to writing it because they have been educated in Croatian. Zegrean (2012) follows the system in Kovačec (1971 and 1978) and Sârbu (1992) and the original orthography when quoting examples from the literature. We follow this general practice despite the fact that it can create some inconsistencies.

#### *3.1 Subject Verb Agreement*

The four languages under consideration display Subject Agreement on the Tensed form of the verb. This is generally common in many languages of the world. Dryer (2011a) points out that 437 out of 711 languages have the pronominal subject expressed by affixes on verbs (and therefore not expressed independently). The presence of subject agreement is however not always sufficient to allow for a null subject pronoun in finite clauses, as is the case of French (Kayne 1975) or some Italian dialects (cf. Manzini and Savoia 2005). Istro-Romanian, Romanian, Italian, and Croatian all behave in the same way, displaying a null subject pronoun, as in (1):

- |     |    |                          |                  |
|-----|----|--------------------------|------------------|
| (1) | a. | Cunosc doi mladić        | (Istro-Romanian) |
|     | b. | Cunosc doi tinerei       | (Romanian)       |
|     | c. | Conosco due giovanotti   | (Italian)        |
|     | d. | Znam dva mladića         | (Croatian)       |
|     |    | ‘[I] know two young-men’ |                  |

In the generative tradition (Rizzi 1982; Burzio 1986 for Italian; Dobrovie Sorin 1994 for Romanian), this property is correlated with the possibility of VS order, in particular with unaccusative verbs. Our four languages all confirm this tendency:



- (2) a. Vire ur cârștiân (Istro-Romanian)  
 b. Vine un om (Romanian)  
 c. Arriva un uomo (Italian)  
 d. Dolazi neki čovjek (Croatian)  
 ‘[‘There] comes a man’

Subject agreement, null subject pronouns, and postverbal subjects make Istro-Romanian as perfectly congruent with both Croatian, eastern Romance, and Central Romance.

### 3.2 Auxiliary Selection and Past Participle Agreement

Istro-Romanian is parallel to Romanian in displaying only one auxiliary to express past Tense, namely *have* and no past participle agreement with the subject any verbs. We observe unaccusatives in (3) and transitives in (4):

- (3) a. Kristina a verit (Istro-Romanian)  
 b. Cristina a venit (Romanian)  
 ‘Cristina has arrived’
- (4) a. Lara a poidit paninu (Istro-Romanian)  
 b. Lara a mâncat un sandwich (Romanian)  
 ‘Lara has eaten a sandwich’

In this respect, Istro-Romanian is quite different from Croatian, which behaves like other Slavic languages in forming the past tense with an adjectival form of the verb agreeing with the subject and occurring with the auxiliary *be*, with both unaccusatives (5a) and transitives (5b):

- (5) a. Kristina je stigla  
 Cristina is arrived.F.SG
- b. Lara je pojela sendvič  
 Lara is bought.F.SG. [a] sandwich

The presence of auxiliaries to form past tenses is an innovation of Romance with respect to Latin. It must be noted that all attested medieval Romance languages display the *be/have* alternation that differentiates unaccusative verbs with auxiliary *be* and subject agreement on the past participle, and all other verbs (transitives, ditransitives and intransitives) with auxiliary *have* and no agreement on the past participle. Here we give an Italian example:

- (6) a. Cristina è arrivata  
Cristina.FSG is arrived.FSG
- b. Lara ha mangiato un panino  
Lara.FSG has eaten a sandwich

The gradual loss of the *bel/have* alternation is witnessed in the historical grammar of Spanish. We are not aware whether this has been the case of Romanian as well, whose earliest attestations are however later (from the 16<sup>th</sup> c.). However, Alexandru Nicolae (p.c.) draws our attention on the possibility in modern Romanian to have something morphologically though not semantically parallel to (6a): *Cristina e venită (de ieri)* (lit. C. is come (since yesterday), ‘Cristina has been here since yesterday’). This property is also present in the other Balkan Romance languages (cf. Mišeska-Tomić 2006). We could suppose that this is a residue of an older stage with the same auxiliary alternation we find in the more central Romance languages. In any case, the contact with Croatian does not seem to have had any influence on auxiliary selection in Istro-Romanian.

### 3.3 Verb Movement

Istro-Romanian is quite similar to Croatian in what in generative grammar is called “Long Verb Movement”, which consists in fronting the past participle to the sentence initial position at the left of the auxiliary. We call this phenomenon V-preposing, to make our terminology theoretically agnostic:

- (7) a. Stigla je (Croatian)  
come.FSG. is
- b. Verit-a (Istro-Romanian)  
come-has  
‘She arrived’

V-preposing is obligatory when the sentence only consists in auxiliary and past participle. This is due to the fact that the auxiliary in Croatian (8a) is a clitic in Wackernagel position. The same appears to hold in Istro-Romanian (8b):

- (8) a. \*Je stigla (Croatian)  
be come.FSG
- b. \*A verit (Istro-Romanian)  
has come

Italian behaves in the opposite way (9a). V-preposing is impossible (9b), unless the past participial form is contrastively focused (9c):

- (9) a. È arrivata (Italian)  
is arrived.F.SG  
b. \*Arrivat(a) è  
c. ARRIVATA è (NON partita)  
arrived.F.SG [she] is (not departed.F.SG)

Today's Romanian (10a) is more similar to Italian, but V-preposing (10b) can be found in the literary register and in some tenses<sup>5</sup> (cf. Dobrovie Sorin 1994; Rivero 1994). Note that it was more common in Old Romanian as witnessed by (10c):

- (10) a. A venit (Romanian)  
has come  
b. #Venit-a (Istro-Romanian)  
Come=has  
c. Cumparat-a den voivodesei den Țara Leșescu  
[He/She] bought=has from voivod.F.SG.Gen from Țara Leșescu  
(*CI Inventar Bozen*, 1594, from Chivu *et al.* 1979)

This piece of data suggests that V-preposing in Istro-Romanian is the heritage of a common Balkan Romance origin. It is, in fact, also present in Megleno-Romanian (Mišeska-Tomić 2006: 377-78).

It is well known that isolated languages have a conservative tendency. The contact with Croatian which has a similar, though not identical phenomenon (cf. Čavar and Wilder 1994), has certainly played a major role in the preservation of this property which is now almost lost in modern Romanian. This hypothesis is confirmed by the fact that the conditions on the optional vs. obligatory nature of V-preposing are not exactly the same in Istro-Romanian and Croatian.

For example, in Croatian the auxiliary must immediately follow a preverbal DP subject, as in (11). This is not the case in Istro-Romanian (12):

<sup>5</sup> Oana Savescu makes us notice that verb preposing is more natural with conditional verbs that are formed by a special form of the auxiliary *have* plus the infinitive form of the verb. So that a sentence like (i) has a grammatical counterpart with verb preposing in (ii):

- (i) eu l-as mânca  
I CL would eat  
(ii) Mânca-l-as  
eat CL would  
'I would eat it'

- (11) a. Redento je čitao knjigu (Croatian)  
 Redento is read book.Acc  
 'Redento read the book'
- b. \*Redento čitao-je knjigu (Istro-Romanian)  
 Redento read-is book. Acc
- (12) a. Redento čiteit-a libru (Istro-Romanian)  
 Redento read=has book.the  
 'Redento read the book'
- b. Vaca durmit-a tota nopta  
 cow-the slept=has all night-the  
 'The cow slept all night long'
- c. Kristina verit-a ier  
 Kristina come=has yesterday  
 'Kristina has arrived yesterday'

Our elicited corpus does not contain the Istro-Romanian counterpart of (11a); but the ungrammaticality of (13b) suggests that the auxiliary cannot encliticize onto a full DP subject:

- (13) a. Dejan mes-a ân beserica (Istro-Romanian)  
 Dejan gone-has in church
- b. \*Dejan a mes ân beserica  
 Dejan has gone in church  
 'Dejan went to church'

But an auxiliary can encliticize on non-verbal constituents, as we observe with the Time adverb 'yesterday' in (14a), and the complementizer 'that' in (14b):

- (14) a. Ier-a Redento čuda čiteit (Istro-Romanian)  
 yesterday=has Redento much read  
 'Yesterday Redento read a lot'
- b. Se ganę k-a Redento čuda čiteit  
 SE says that=has Redento much read  
 'It is said that Redento read a lot'

Example (14) also shows that the enclitic auxiliary can move above the subject in Istro-Romanian: a property that is not shared by Romanian and is instead found in Croatian.

Furthermore, the clitic cluster formed by the auxiliary and the clitic negation (15a) or a clitic object (15b) can stand alone and does not need a host:

- (15) a. Nu-a nič čiteit (Istro-Romanian)  
           not=has nothing read  
           ‘[He/She] read nothing’  
       b. L-am vezut  
           [I] him.AccCL-have1SG seen  
           ‘I saw him’

It is important to note the asymmetry between third person on the one hand and first/second person on the other. It is apparent that first and second person pronouns in (16) can serve as hosts for the clitic auxiliary (or as part of the cluster containing the clitic auxiliary) while third person pronouns cannot, parallel to full DPs, as observed in (13) above:

- (16) Indicative present perfect tense: auxiliary + participle (of *fini* ‘to finish’)

SG	1	io-m finit	PL	noi am finit
	2	tu-i finit		voi aț finit
	3	je finit-a (*je a finit)		celj finit-a (*celj a finit)

A frequency adverb like *vet* ‘already’, which is clearly borrowed from Croatian, gives unexpected results in (17a). It allows for the auxiliary to occur immediately after a DP subject, differently from (13b) above, suggesting that in some cases, the auxiliary can be proclitic onto the following word (the adverb in this case). Furthermore, in (17a) the adverb can optionally appear after the past participle, which mildly contrasts with Croatian that disprefers the adverb after the past participle (17b), and sharply contrasts with Romanian that cannot split the auxiliary and the past participle in (17c):

- (17) a. Kristina a {vet} verit {vet} (Istro-Romanian)  
       b. Kristina je {već} stigla {već} (Croatian)  
       c. Cristina a {\*deja} venit {deja} (Romanian)  
           ‘Christine has (already) arrived (already)’

Romanian only has a restricted number of weak adverbs that may appear between auxiliary and past participle: *mai*, *tot*, *cam* (cf. Dobrovie-Sorin 1994). Istro-Romanian may have extended this possibility (or maintain a possibility available in an early stage of the language) in contact with Croatian.

To conclude, as regards the position of the auxiliary and the past participle, it seems that Istro-Romanian preserves a property of Old Romanian that is, at least superficially, shared with Croatian, namely V-preposing. Furthermore, it has not developed a strict adjacency requirement of the Aux–V order, that is found in Modern Romanian with most adverbs, probably in contact with Serbo-Croatian.

### 3.4 Clitic Clusters

Istro-Romanian, parallel to Romanian, presents a positional difference of masculine vs. feminine third person accusative clitic pronouns, in that the feminine clitic appears more embedded in the verbal cluster: in compound tenses, masculine *l-* is proclitic to the auxiliary (18a), while feminine singular *vo* is found between the auxiliary and the past participle (18b). We are not sure whether *vo* is enclitic on the auxiliary or proclitic on the lexical verb, but we note in (18c) that it is certainly not enclitic to a lexical verb:

- (18) (Istro-Romanian)
- |    |   |   |
|----|---|---|
| a. | Nu <b>l-</b> am vezut<br>[I] not him.CL have.1SG seen<br>'I haven't seen him' | Neg > <b>CL<sub>masc</sub></b> > AuxCL > PastPart |
| b. | Nu-am <b>vo</b> vezut<br>[I] not have.1SG her.CL seen<br>'I haven't seen her' | Neg > AuxCL > <b>CL<sub>fem</sub></b> > PastPart  |
| c. | *Nu-am vezut <b>vo</b>  | Neg > AuxCL > PastPart > <b>CL<sub>fem</sub></b>  |

The cases in (22b-c) sharply contrast with their Romanian counterparts (23b-c), (cf. Dobrovie-Sorin 1994):

- (19) (Romanian)
- |    |   |   |
|----|---|---|
| a. | Nu <b>l-</b> am văzut<br>I not him.CL have.1Sg seen<br>'I haven't seen him' | Neg > <b>CL<sub>masc</sub></b> > AuxCL > PastPart |
| b. | *Nu-am <b>o</b> văzut<br>I not have.1SG her.CL seen                         | Neg > AuxCL > <b>CL<sub>fem</sub></b> > PastPart  |
| c. | Nu-am văzut- <b>o</b><br>I not have.1SG seen-her.CL<br>'I haven't seen her' | Neg > AuxCL > PastPart > <b>CL<sub>fem</sub></b>  |

This difference can be related to the possibility in Istro-Romanian vs. the impossibility in Romanian to separate the auxiliary from the past participle (cf. 17a above). Once again, we find a property related to the apparent attraction of the auxiliary towards the initial part of the clause, which makes Istro-Romanian similar to Croatian.

Istro-Romanian is very similar to other languages in displaying the order dative > accusative in the clitic cluster. As may be expected given the properties observed above, the cluster can appear sandwiched between a preposed past participle and the auxiliary, or it can serve as the host of the second position auxiliary.

- (20) (Istro-Romanian)
- a. Dăt -Ți -l -am. PastPart > **CLDat** > **CLmascAcc** > AuxCL  
 given you.Dat it.CL have.1SG
- b. Ți l -am dăt. **CLDat** > **CLmascAcc** > AuxCL > PastPart  
 you.Dat it.CL have.1SG given  
 'I gave it to you'

The differences and similarities with Romanian and Croatian are expected. Modern Romanian does not really allow any past participle preposing. For this reason (21a) is very marked, though not ungrammatical:

- (21) (Romanian)
- a. #Datu -Ți -l -am. PastPart > **CLDat** > **CLmascAcc** > AuxCL  
 given you.Dat it.CL have.1SG
- b. Ți l -am dat. **CLDat** > **CLmascAcc** > AuxCL > PastPart  
 you.Dat it.CL have.1SG given  
 'I gave it to you'

On the other hand, since the Croatian auxiliary is in Wackernagel position, the clitic cluster cannot intervene between the fronted past participle and the auxiliary, as in (22a). The auxiliary is the first in the cluster and is followed by the pronouns, which display the dative > accusative order observed in Romance, as in (22b). The other possibility is when the clause is introduced by an adverb, thus the order Aux > dative > accusative is also present in case an adverb is in first clausal position as in (22c):

- (22) (Croatian)
- a. \*Dao/\*Dala ti ga sam  
 given.M/F you.Dat it.CL have.1SG
- b. Dao/Dala sam ti ga  
 given.M/F you. have.1SG Dat it.CL  
 'I gave it to you'
- c. Juče sam ti ga dao/dala  
 Yesterday have.1SG you.Dat it.CL given.M/F  
 'Yesterday I gave it to you'

From the two properties of clitic clusters observed above, the question arises as to how the feminine clitic is positioned in the case of V-preposing in Istro-Romanian. A minimal pair with the masculine is given in (23)-(24):

- (23) a. Redento l-a vezut  
Redento he.Acc.CL = has seen  
b. Redento vezut-l-a  
Redento seen he.Acc.CL =has
- (24) a. Redento {vo} vezut-a {vo}  
Redento {her.Acc.CL} seen = has {her.Acc.CL}  
'Redento saw him/her'  
b. \* Redento vezut(-v)o-a (also cf.) \*Nu-am vezut (v)o

The contrast in (23)-(24) shows differently from Romanian, in Istro-Romanian the feminine clitic can be procliticized onto the fronted past participle as in (24a) and is never enclitic on the finite verb (24b), confirming what was shown in (18c).

### 3.5 Yes/No Questions and Short Answers

Istro-Romanian presents an interrogative particle *če* which is obligatory in wide scope *yes/no*-questions (25a). If it is missing, part of the question is presupposed. The interpretation of (25b) can be paraphrased as '(How about) Kristina, did she come?' while the paraphrase of (25c) is 'Was it Kristina who arrived?':

- (25) a. Če-a verit Kristina? (wide scope question)  
PRT-has come Kristina  
b. (\*Če) Kristina a verit? (narrow scope question on the subject)  
c. (\*Če) Verit-a Kristina? (narrow scope question on the verb)  
'Did Kristina come?'

*Če* occupies the first position in the clause, and as such is the host of the clitic cluster in second position. Compare the ungrammatical version of (25b-c) with the grammatical examples in (26):

- (26) a. Če-ai fost în čine?  
PRT have2SG been in cinema  
'Have you been to the cinema?'  
b. Če-l vezi?  
PRT him.CL see.2SG  
'Do you see him?'  
c. Če nu vo vezi?  
PRT her.CL see.2SG  
'Don't you see her?'



In this respect, Istro-Romanian is completely parallel to Chakavian, which presents a similar particle *ča* (27). But it is not dissimilar to some Italian dialects, like for example Anconetano in (27b). This shows that the phenomenon is also present in Romance, despite its absence in Romanian (28a-b), which presents the question marker *ce* only in the *ce face* construction:

- (27) a. Ča je prišla Kristina? (Chakavian)  
       b. Che è venuta Cristina? (Anconetano)  
           PRT is come Cristina
- (28) a. Cristina a venit? (Romanian)  
       Cristina has come  
       b. A venit Cristina?  
           Has come Cristina  
       c. Ce face Cristina, a venit? (Substandard Romanian)  
           What does Cristina, has come  
           ‘Has Cristina arrived?’

A contact-induced linguistic property such as the formation of an interrogative marker is easily integrated in a system that is able to accommodate it, as the Romance system, with a rich left peripheral expansion of the clause. This is in all cases a common property of the languages of the world, as shown by Dryer (2011b) who reports that 584 among 954 languages display an interrogative marker to mark Yes/No Questions.

A striking property that distinguishes Istro-Romanian from all other Romance languages, is the possibility of VP deletion, at least as regards short answers. To a question like ‘Did you sleep?’ with an intransitive verb, and the first/second person alternation in the answer we can obtain two possible answers as in (29):

- (29) Če -ai durmit ier? (Romanian)  
       PRT have.2SG slept yesterday
- a. Durmit -am  
    slept have.1SG
- b. Ăm.  
    have.1SG  
    ‘Did you sleep yesterday? Yes, I slept. / I did.’

As regards the (quasi-)clitic nature of the auxiliary observed in 3.2 above, it is interesting to note that while the first person auxiliary, formed by a vowel and a consonant can stand alone (29b), there is variation as regards the second person, formed by a diphthong, which requires for some speakers a preceding consonant *v-*, which we could not relate to any element and for some speakers

is a condition for the acceptability of the short answer (30a). The consonant disappears if the auxiliary is clustered with the negation (30b):

- (30) Če m-am ponešeit bire âz?  
PRT REFL.1SG have.1SG behaved well today
- a. %(V-) Ai.  
have.2SG
- b. N-ai.  
not-have.2SG  
'Have I behaved well today? You have / You haven't'

The weakest form of the auxiliary is the third person, which either requires V-preposing (31a), or is obligatorily spelled out with the consonant (31b-c):

- (31) Če mes-a az Goran ân Trst?  
PRT gone have.3SG today G. inTriest
- a. Mes -a.  
gone have.3SG
- b. \*A.  
have.3SG
- c. V-a.  
CONS-have.3SG  
'Did Goran go to Triest today? Yes, he went'

This property is only shared by Croatian and is not present in Romanian, or in any Romance variety that we know of:

- (32) a. Da, imam. (Croatian)  
Yes, [I] have
- b. Ne, nemam  
No, [I] not-have
- c. Da, hoću  
Yes, [I] will
- d. Ne, neću  
No, [I] not-will

In this subsection we have observed two contact induced features of Istro-Romanian that distinguish it from Romanian, namely the presence of an interrogative marker in Yes/No questions and the possibility to have VP-deletion in short answers.

### 3.6 A Provisional Protocol for Istro-Romanian

In this Section we have documented some properties of Istro-Romanian clausal syntax in a comparative perspective with Romanian, Croatian and Italian. They can be clustered under 6 headings, as in the following chart:

(33)	Istro-Romanian	Romanian	Croatian	Italian
Subject-Verb Agreement	+	+	+	+
SV/VS order	+	+	+	+
null subj. pro in finite clauses	+	+	+	+
Aux in past.T / perf.Asp:	+	+	+	+
<i>have</i> - PastPart agr	+	+	-	+
<i>be</i> + PastPart agr	-	-	+	+
Clitics	+	+	+	+
subject pronouns	-	-	-	-
object pronouns	+	+	+	+
auxiliary	+	+	+	-
negation	+	+	+	+
Aux in Wackernagel pos.	+/-	(+)/-	+	-
3rd.P Subject host	-	#	+	#
1st/2nd. P. Subject host	+	#	+	+
Aux Adv V	+	+/-	+	+
V-preposing	+	(+)/-	+	-
Clitic clusters with Aux	+	+	+	+
Acc.M/F in diff. pos.	+	+	-	-
PastPart Cl Aux	+	(-)	-	(-)
CL Aux PastPart	+	+	-	+
PastPart Aux Cl	-	-	+	-
overt interrog. marker	+	-	+	(+)/-
short answers	+	-	+	-

The clusters of properties proposed in (32) are necessarily tentative and incomplete, due to limitations of space and more importantly of in-depth knowledge of this language. The aim of the paper has been to propose the protocol methodology, designed to be quite flexible to include future advances of comparative linguistics.

For example, a more in-depth research on VP-deletion could relate the possibility of short answers to V-preposing. A deeper knowledge of clitic clusters could relate the descriptive clusters provided in 5 with the Wackernagel vs. T-position of the auxiliary. A better way of capturing person restrictions is necessary, as the ban of encliticization of the auxiliary on the third person subject may depend on properties independent of the Wackernagel position of the auxiliary. Of course, many properties have been left

out, as for example the presence vs. absence of infinitival clauses, which is a major Balkan feature.

The protocol methodology can easily be extended to any aspect of language, including lexicon, phonology, but also pragmatics, discourse features, etc. We have started with syntax for two major reasons, apart from the fact that we are syntacticians: first because syntax is less considered in endangered language methodology, and second because the type of comparison done here is directly based on the new comparative methodology of the principles and parameters framework.

#### 4. Conclusion

Despite its incompleteness, the protocol in (32) can be a good start in writing a grammar of Istro-Romanian for documentation, revitalization, standardization, as well as for educational purposes. Linguistic metacompetence must be made available to all actors in the process of regaining power on the language and approaching the issue of the historical origin in a more objective perspective.

Despite only taking into account a few properties, the Protocols presented above permit to obtain an inclusive perspective of Istro-Romanian across at least four dimensions. First of all, the universal dimension: the possibility in a language to omit the subject, related to richness in verbal morphology, and to find a DP subject in post-verbal position is shared by the languages observed and is rather common worldwide. Second, the inter-Romance dimension: the lack of the *bel/have* auxiliary alternation collocates Istro-Romanian in the Balkan-Romance group, distinguishing it from central Romance (Italian). Third, the contact with Croatian has strengthened properties that are going to be lost in Romanian, as is the case of V-preposing (marginally present in Romanian but more frequent in old Romanian and Megleno-Romanian). It has created an innovation, such as the interrogative marker (that is compatible with the rich clausal left periphery of Romance varieties as witnessed by Italian dialects, and is a wide-spread property among the languages of the world). It has also created two phenomena that (to our knowledge) are not found in any Romance languages, namely short answers and VP deletion and the (quasi) Wackernagel position of auxiliaries. Finally, the different position of accusative masculine vs. feminine clitics is a peculiarity of Romanian that is found with micro-parametric differences in Istro-Romanian, which has a different behavior of the auxiliary (also due to contact with Croatian).

From this brief overview, two major aspects of the language emerge: the historical origin of Istro-Romanian as belonging to Balkan-Romance and the strong contact with Croatian, a language that in modern time is the (other) first language of the speakers. These dimensions are two equally important sides of the same precious coin, in that they both concur in making the language

not as an isolated individual species, but as the result of language contact and change. If language is one of the fundamentals of cultural identity, disseminating metalinguistic awareness of these aspects can help the community of speakers interact with the research community of linguists to support language documentation, plan the means of dissemination of metalinguistic competence in education, and ultimately construct inclusive cultural identity that is not in contrast with the national Croatian identity of the speakers but is also not in contrast with the eastern Romance nature of the language and at the same time could easily overlap with identities construed outside Croatia.

In this perspective, we hope, Protocol Linguistics can contribute to enhancing inclusive cultural identity, in the perspective of one of the Horizon 2020 goals; namely, to create “inclusive, innovative and secure societies”<sup>6</sup>.

### References

- Brandi, Luciana, and Patrizia Cordin. 1989. “Two Italian Dialects and the Null Subject Parameter.” In *The Null Subject Parameter*, ed. by Osvaldo Jaeggli, and Kenneth J. Safir, 111-142. Dordrecht: Kluwer.
- Burzio, Luigi. 1986. *Italian Syntax: A Government and Binding Approach*. Dordrecht: Springer.
- Čavar, Damir, and Christopher Wilder. 1994. “Word Order Variation, Verb Movement, and Economy Principles.” *Studia Linguistica* 48, 1: 46-86.
- Chivu, Gheorghe, Magdalena Georgescu, Magdalena Ioniță, Alexandru Mareș, and Alexandra Roman-Moraru (eds). 1979. *Documente și însemnări românești din secolul al XVI-lea*, București: Editura Academiei.
- Chomsky, Noam. 1981. *Lectures on Government and Binding*, Dordrecht: Foris.
- Dobrovie Sorin, Carmen. 1994. *The Syntax of Romanian*. Berlin: Mouton.
- Dryer, Matthew S. 2011a. “Expression of Pronominal Subjects.” In *The World Atlas of Language Structures Online*, ed. by Matthew S. Dryer, and Martin Haspelmath. Munich: Max Planck Digital Library, Chapter 101. <<http://wals.info/chapter/101>> (09/2015).
- Dryer, Matthew S. 2011b. “Polar Questions.” In *The World Atlas of Language Structures Online*, ed. by Matthew S. Dryer, and Martin Haspelmath. Munich: Max Planck Digital Library, Chapter 116. <<http://wals.info/chapter/116>> (09/2015).
- Dryer, Matthew S. 2013. “Order of Subject and Verb.” In *The World Atlas of Language Structures Online*, ed. by Matthew S. Dryer, and Martin Haspelmath. Munich: Max Planck Digital Library, Chapter 82. <<http://wals.info/chapter/82>> (09/2015).
- Edwards, John. 2009. *Language and Identity. An Introduction*. Cambridge: Cambridge UP.

<sup>6</sup> <[http://ec.europa.eu/research/horizon2020/pdf/proposals/com%282011%29\\_808\\_final.pdf](http://ec.europa.eu/research/horizon2020/pdf/proposals/com%282011%29_808_final.pdf)> (09/2015).

- Joseph, Brian. 1993. *Romanian and the Balkans: Some Comparative Perspectives*. <<http://www.ling.ohio-state.edu/~bjoseph/publications/1999roma.pdf>> (09/2015).
- Kovačec, August. 2009. "Some Facts about Istro-Romanian", Handout of the Talk given at The Days of the Vlaški or Žejanski Language, July 10-12, Žejane. Croatia.
- Manzini, M. Rita, and Leonardo M. Savoia. 2005. *I dialetti Italiani e Romanci. Morfosintassi Generativa*. Alessandria: Edizioni dell'Orso.
- Mišeska-Tomić, Olga. 2006. *Balkan Sprachbund. Morphosyntactic Features*. Dordrecht: Springer.
- Pușcariu, Sextil. 1926. *Studii istroromâne* 2. București: Cultura Națională.
- Rivero, Maria Luisa. 1994. "Clause Structure, V-Movement and the Languages of the Balkans." *Natural Language and Linguistic Theory* 12, 1: 63-120.
- Rizzi, Luigi. 1982. *Issues in Italian Syntax*. Dordrecht: Foris.
- Vrzic, Zvezdana. 2009. "How to write Vlaški and Žejanski?", Handout of the Talk given at The Days of the Vlaški and Žejanski Language, July 10-12, Žejane. Croatia.
- Zegrean, Iulia. 2012. *Balkan Romance. Aspects of the Syntax of Istroromanian*. PhD Diss. Ca' Foscari University of Venice.