



BRILL



brill.com/jlc

Contact-Induced Complexification in the Gender System of Istro-Romanian

Michele Loporcaro

Full Professor of Romance Linguistics, Institute of Romance Studies, Faculty of Arts and Social Sciences, University of Zurich, Zurich, Switzerland

loporcar@rom.uzh.ch

Francesco Gardani

Professor of Romance Linguistics, Institute of Romance Studies, Faculty of Arts and Social Sciences, University of Zurich, Zurich, Switzerland

francesco.gardani@uzh.ch

Alberto Giudici

Assistant in Romance Linguistics, Institute of Romance Studies, Faculty of Arts and Social Sciences, University of Zurich, Zurich, Switzerland

alberto.giudici@uzh.ch

Abstract

The paper provides the first description of the borrowing of Croatian collective numerals into Northern Istro-Romanian and explores the consequences of this borrowing for the morphosyntax of the recipient language. It argues that the collective numerals under examination, which are specified as nominative plural feminine in the Slavic model, took on a different structural specification in the Romance replica, in a way that led to a restructuring of the morphosyntactic system, introducing (sub)gender overdifferentiation on just two agreement targets and, thereby, a complexification in this area of grammar. The illustration of this change is placed against the background of the other contact-induced changes that grammatical gender has undergone in Istro-Romanian during the 20th century, which have led to the borrowing of two dedicated forms in distinct inflectional cells and the rise of two separate defective gender values, each the replica of one number value of the Slavic neuter.

Keywords

borrowing – contact-induced complexification – gender – number – numerals – overdifferentiation – Daco-Romance – Croatian

1 Introduction*

One of the probably most widely accepted claims made in the language contact literature is that contact leads to the simplification of grammar. The basic assumption behind this claim is that the mixing of linguistic systems produces less marked structures and levels out irregularities towards “a kind of common core-grammar” (Mühlhäusler, 1980: 28; see also Givón, 1979; Bickerton, 1981). This simplification hypothesis, despite a few counterexamples (Thomason and Kaufman, 1988: 29; Vanhove, 2001; Aikhenvald, 2002; 2003; Adamou, 2013; de Groot, 2008; Melissaropoulou, 2017; Loporcaro, 2018: 51, 291), still dominates research not only in creole languages (McWhorter, 2001; 2007), but also in so-called regular contact-induced change (Kusters, 2003; Trudgill, 2009; 2011).

In the present paper, we describe a case of contact-induced grammatical complexification involving Istro-Romanian, a heavily endangered Romance language spoken by a few dozen speakers, all bilinguals with Croatian, in two areas of Istria as well as by a few hundred speakers around the world as a heritage language. After providing background information on the language (Section 2) and on numeral borrowing cross-linguistically and in Slavic-Romance contact (Section 3), we will address Istro-Romanian numerals, showing that the borrowing process has concerned not only ordinal and cardinal numbers (a fact that has long been described, Section 4), but also adnominal numerical quantifiers (Section 5). Here, borrowing has made possible numerical quantification with *pluralia tantum* nouns – with which Romance languages often resort to alternative strategies (the “classifier solution”, Corbett, 2019: 93f.; see the examples in (6)–(7) below) – and at the same time the signalling of gender/number agreement with such head nouns on some numerical quantifiers, in a way that

* Whenever unreferenced, the examples provided stem from field recordings which are stored at the Phonogram Archives of the University of Zurich. Glosses follow the Leipzig Glossing Rules: for simplicity, case specification is omitted in IR clauses, where nominal forms are always given in the nominative/accusative case. In addition, we use the following abbreviations: BCS = Bosnian-Croatian-Serbian, (N/S)IR = (Northern/Southern) Istro-Romanian, PT = *plurale/pluralia tantum*. In grammaticality judgements, * = ungrammatical, ? = marginally acceptable, % = acceptable only for some informants. For academic purposes ML must be held responsible for Sections 5–7, 8.1–8.2, 8.4–8.5, and 9, FG for Sections 1 and 3, AG for Sections 2 and 4, FG and AG jointly for Section 8.3.

deviates from what is found elsewhere on agreement targets in Istro-Romanian. We will argue that this resulted in morphosyntactic complexification. First, the morphosyntactic system prior to the borrowing is described in Sections 6–7, showing that contact with Croatian had already made the inherited gender system more complex, triggering the borrowing of two additional gender values in two successive steps. Then, against this background, in Section 8 we argue that the borrowing of the numerical quantifiers at issue has led to the rise of gender overdifferentiation (Corbett, 1991: 168f.) on just a few agreement targets (lower numerals). While gender overdifferentiation on lower numerals has been described before for several languages, including some Romance varieties, its rise through borrowing never has. Since this borrowing process resulted in a net increase in complexity of the gender/number agreement rules, this case study adds to the series of contact-induced changes which bring about complexification, rather than simplification, in the grammatical subsystem involved (cf. e.g., Vanhove, 2001; Aikhenvald, 2003; Adamou, 2013; Melissaropoulou, 2017; Loporcaro, 2018: 51, 291; Meakins and Wilmoth, 2020).

2 Istro-Romanian

Istro-Romanian (henceforth IR) is one of the four branches of Daco-Romance.¹ It is spoken today by a tiny number of speakers (about 100, most of them over 50 years old) in north-eastern Istria (see Map 1). It divides in two mutually comprehensible, yet clearly distinct varieties, which have been isolated from each other for centuries since the late Middle Ages and developed divergent innovations in both lexicon and grammar:² Northern Istro-Romanian

1 There is an issue about terminology here. While most authors call Istro-Romanian one of the four dialects of Romanian (see e.g., Tagliavini, 1972: 356–364), linguists from the local community (e.g., Vrzić and Doričić, 2014: 105) prefer subsuming Žejanski directly under a superordinate classificatory unit ‘Eastern Romance’.

2 The list of diverging grammatical properties includes various differences in verb inflection, e.g., SIR *-rno* 1pl restrictive future employed only in conditional clauses *flârno* (Puscariu, 1926: 185) vs. NIR *aflârem* ← *aflâ* ‘find’ (Kovačec, 1971: 143; see also Hurren, 1999: 101); the loss of the imperfect tense in NIR versus its preservation in SIR, where, however, it is restricted to the aspectual function of continuous (Hurren, 1969: 89).

- (i) a. *in špitól am muŋkót žir svka zi* (NIR)
 b. *in špitól muŋkójam žir svka zi* (SIR)
 ‘while at the hospital, I used to eat fruit every day’

In the lexicon too, several differences exist, often due to the different intensity of contact with different languages in SIR vs. NIR: for instance, for ‘newspaper’ NIR uses the Croatian loan

(henceforth NIR), spoken only in the village of Žejane (IR *Jeiān*, Italian *Seiano*, in the municipality of Matulji, Primorje-Gorski Kotar district), and Southern Istro-Romanian (henceforth SIR), spoken in a cluster of villages lying some 22 km to the SSW as the crow flies, but at least 40 km on foot, being separated from NIR by the Učka/Monte Maggiore massif.³ For SIR, the data cited in this



Map 1. Istro-Romanian in Istria (after Loporcaro, 2018: 293, with modifications)

■ = Istro-Romanian; ◆ = Croatian; ● = Italo-Romance and Croatian

novine (plurale tantum), whereas SIR has borrowed *ĝornble* from Italian. A detailed account of these differences is provided by Kovačec's (1998) dictionary and Filipi's (2002) atlas.

3 The villages, all included in the municipality of Kršan (in the Istria district), are those of Brdo (IR *Bārda*, It. *Berdo*), Kostrčani (IR *Costārcian*, It. *Costerciani*), Letaj (IR/It. *Letaj*), Miheli (= IR/

article come from Šušnjevića when not otherwise specified, as well as from nearby Jesenovik.

All IR speakers are bilingual with Croatian (in the standard variety and the Čakavian varieties). As a result of centuries-long total language contact, the structure of IR has been massively reshaped (see Kovačec, 1963; 1966; 1968; 1971; Filipi, 2002; Sala, 2013: 218–225; Vrzić and Doričić, 2014). In the phonology, consonants with a secondary palatal articulation lost it (the contrast does not occur in Croatian), which impacted on inflectional morphology, since in Romanian palatalization induced by final *-i/* marks the plural in many nominal paradigms, whose singular/plural forms became homophonous in IR: compare Ro. *lup* ‘wolf’, pl. *lup* with IR *lup* ‘wolf=wives’ (Kovačec, 1998: 108).⁴ The syntax of IR is basically that of Croatian, including its relatively free word order (contrary to Romanian) as well as specific rules such as those affecting the placement of clitic auxiliaries (e.g., *vlbt=am* ‘I have taken’, vs. Ro. *am luat*; see the examples in (14a,c) below). In the lexicon, extensive borrowing resulted in replacement even in core lexical domains: Vrzić and Doričić (2014) describe its increase over time for body parts. As a consequence, often whole IR sentences consist solely of Croatian lexical material “sans en changer autre chose que les morphemes grammaticaux” [without changing anything else but grammatical morphemes] (Kovačec, 1968: 81). Even here, Croatian has impacted, as witnessed by IR being possibly the sole Romance language in which the inherited first conjugation (Lat. *LIGARE* > *levó* ‘to tie’) has become unproductive, while new verb lexemes are formed with Slavic suffixes (Kovačec, 1971: 131f.): e.g., *čiravéi* ‘to have dinner’, derived with the suffix *-av-* plus a non-etymological inflectional ending from the Romance base (cp. *čira* ‘dinner’ < Lat. *CĒNAM*).

IR speakers are not singled out by a specific ethnic/linguistic identity and perceive themselves as homogeneous to the Croatian environment, a circumstance which favours assimilation. In this ecological setting, generalized bilingualism and the steep increase in mobility over the past few decades triggered a language shift which is now approaching its final stage: IR nowadays does not appear to have fluent native speakers below 40 years of age and is not being passed on to the next generations.⁵ During fieldwork in Istria in 2017–2018,

It. *Micheli*), Nova Vas (IR *Noselo/Nosela*, It. *Villanova*), Šušnjevića (IR *Suseni*, It. *Susnevizza*), Jesenovik (IR *Sukodru*, It. *Frassineto*), and Zankovci (IR *Zancovti*, It. *Zancovzi*) (see the list in Filipi, 2002: 31).

4 For some lexemes, the distinction was restored applying the *-ure* suffix originally restricted to neuters, as in IR *lúpure* ‘wives’ competing with the unmarked plural *lup* (Kovačec, 1966: 64; see examples in context in (36a-b)).

5 Language shift is rampant in the area, so that the *Ethnologue* classification as “shifting” (EGIDS level 7: cf. <https://www.ethnologue.com/cloud/ruo>) is more than justified. Given this, it is obvious that higher figures are indicated in the literature as one climbs back in time.

we had a chance to interview a dozen IR speakers: the results brought to light some interesting facts that had gone unnoticed in the previous literature.

3 Borrowed Numerals and Mixed Numeral Systems in Slavic-Romance Contact and Beyond

The borrowing of numerals is a phenomenon that recurs cross-linguistically (cf. e.g., the examples gathered by Matras, 2009: 201–203, from different language families, ranging from Indonesian to Vietnamese to Tasawaq; or those provided in the articles covering several languages in Haspelmath and Tadmor, 2009), such that students of language contact have drawn generalizations concerning the borrowability of numerals. In the following scale, proposed by Matras (2007: 61) for the borrowability of word classes, the place of numerals is rather low (emphasis added):

word classes affected by contact (Matras, 2007: 61): nouns, conjunctions > verbs > discourse markers > adjectives > interjections > adverbs > other particles, adpositions > **numerals** > pronouns > derivational affixes > inflectional affixes

While numerals are generally rather resistant to borrowing – a fact that may depend on the “assumption that all languages have some form of quantification” according to Matras (2007: 50) – there seems to be a difference in degree of borrowability between low and high numerals, in such a way that Thomason and Kaufman (1988: 74) set the borrowing of low numerals at level 3 of their scale, the one that requires, “more intense contact”. In this respect, Matras observes the following:

It appears, then, that while higher and more abstract numerals are vulnerable to borrowing due to their association with formal contexts of use, and numerals in general may become borrowing-prone through intensification

From the discussion in Combi (1859: 108f.) and Ascoli (1861: 48f.), it results that the overall demographic size of IRS was over 3000 about the half of the 19th century, while one century later, Tagliavini (1972: 364; first edn. 1949) and Kovačec (1971: 23) reported some 1500 speakers (NIR + SIR). More recently, Filipi (2002: 53) estimates some 90 speakers of SIR and some 80 of NIR, while Vrzić and Doričić (2014: 107) reckon 120 fluent speakers of NIR are left, all over 50 – a steep decrease, which is due partly to depopulation, partly to language shift to Croatian by the speakers still residing in the villages. The truth of the matter is that data are shaky and uncertain: in the same year, Vuletić (2014: 191 n. 9) indicates 53 NIR speakers (out of the 134 inhabitants of Žejane), based on data from the <http://www.vlaski-zejanski.com/> website, provided by the first author of the previous paper (Z. Vrzić).

of economic activity in the (potentially) donor language, the proximity constraint protects ‘salient’ numerals, primarily those below ‘ten’ or ‘five’, but sometimes also ‘ten’ and even ‘hundred’. With the latter two exceptions, and the exception of ‘zero’ whose affinity is toward the formal-abstract numerals, most attested cases add up to support an implicational hierarchy of numeral borrowing: higher > lower numerals. (MATRAS, 2009: 202)

Cross-linguistically well-known cases include e.g., Japanese, where “With a few lexical exceptions, the native system is now used only up to ‘10’; above ‘10’, even those counters which prefer the native numerals must use the Chinese set” (Martin, 2004: 767). Mixed numeral systems have developed also in language contact between Romance and Slavic. For Molise Slavic, Breu (2013) describes the progression in real and apparent time of numerals borrowed from the adjacent Italo-Romance dialects, with the elderly generation using two alternative forms between ‘5’ (*pet/čing*) and ‘10’ (*desat/dijač*), and only loan numerals from ‘11’ on, while the younger generation has generalized the loans from *sěj* ‘6’ on and no longer uses native *šest* ‘6’ etc. In our case study, the borrowing direction is the other way round, from Slavic into Romance.

4 The Impact of Language Contact on Numerals in Daco-Romanian and in IR

Daco-Romance offers interesting material in this area even outside IR. As is well known, Romanian borrowed *sută* ‘100’ from Old Slavic *sŭto*, which has been adapted as feminine like all *o*-ending neuters among ancient loanwords from Slavic (Mihăilă, 1960; Petrovici, 1962; Buchi, 2006: 75f.; Livescu, 2008: 2648). In addition, Romanian calqued all numerals from ‘11’ on, except inherited *mie* ‘1000’: *unsprezece/nouăsprezece* ‘11/19’ = OBlg. *jedinŭ/devęti na desęte, doizeci* = OBlg. *dŭva desęti* ‘20’, etc. (cf. e.g., Schulte, 2009: 248). Istro-Romanian, which has been under contact pressure for centuries, goes much further (see Pușcariu, 1926: 153f.; Kovačec, 1966: 65f.; 1971: 117; 1998: 284; Dahmen and Kramer, 1976: 88; Frățiță and Bărdășan, 2010: 39; Sala, 2013: 220). As the data in (1) show, cardinal numerals from the Daco-Romance native stock are preserved up to 7 in both branches, while beyond that point, SIR replaces 8 and 9 and NIR 9 and 10 with Croatian loanwords:⁶

6 IR data collected in our fieldwork sessions are reported in a simplified IPA transcription: primary stress is marked as \acute{V} (not [‘V]) and only on non-paroxytonic words; palatal consonants are transcribed [š ž č ģ] instead of [ʃ ʒ tʃ dʒ]. Please note that due to typographical reasons IPA [a] and [æ] appear as [a] and [æ] when italicized. Data by other scholars are given in the original orthography. We use the standard orthography for Croatian dialect data (Čakavian).

(1) Cardinal numerals in the languages under discussion: Ro(manian), Lat(in), Cro(atian)

	1 M/F/N	2 M/F/N	3 M,F/N	4	5	6
NIR	<i>ur/ura</i>	<i>doi/do</i>	<i>trei</i>	<i>patru</i>	<i>činč</i>	<i>șvse</i>
SIR	<i>ur/ure</i>	<i>doi/do</i>	<i>trei</i>	<i>patru</i>	<i>činč</i>	<i>șvse</i>
Ro	<i>unu/una</i>	<i>doi/două</i>	<i>trei</i>	<i>patru</i>	<i>cinci</i>	<i>șase</i>
Lat	<i>ūnus/ūna/ ūnum</i>	<i>duo/duae/ duo</i>	<i>trēs/tria</i>	<i>quattuor</i>	<i>quīnque</i>	<i>sex</i>
Cro	<i>jedan/jedna/ jedno</i>	<i>dva/dvije/dva</i>	<i>tri</i>	<i>četiri</i>	<i>pet</i>	<i>šest</i>
	7	8	9	10		
	<i>šppte</i>	<i>opt</i>	<i>dēvet</i>	<i>deset</i>		
	<i>šppte</i>	<i>osam</i>	<i>dēvet</i>	<i>zetse</i>		
	<i>șapte</i>	<i>opt</i>	<i>nouă</i>	<i>zece</i>		
	<i>septem</i>	<i>octō</i>	<i>novem</i>	<i>decem</i>		
	<i>sedam</i>	<i>osam</i>	<i>dēvet</i>	<i>deset</i>		

From 11 on, all numerals (including ‘1000’) are borrowed: for instance, ‘11’ is *jedanăist* (< Čakavian *jedanajst*; cf. standard Croatian *jedanaest*), and *sto* ‘100’ is a secondary borrowing from Slavic, which replaced the older Daco-Romance adapted loanword (*o*) *sută*. In the higher tenths, the multiples of 10 are all borrowed, while units are Romance from 1 to 4 and Slavic from 5 onwards:⁷ for instance, for ‘25’, *dvadeset ši pet* is more frequently used, according to Kovačec (1971: 117), than *dvadeset ši činč*. Climbing back in time, one can follow the spread of Slavic loans as well as other contact-induced phenomena through the sources. For example, Ugo Pellis (cf. Dahmen and Kramer, 1988: 222–224) mentions that in Žejane (at the time, as the whole of Istria, under Italian rule) the Italo-Romance (Venetan) numerals could be used as an alternative, which is no longer the case today. Otherwise, his data match those reported in Pușcariu (1926) and the later sources cited above. On the contrary, Ascoli (1861: 75), based on Combi (1859: 99–139), reports for NIR the Daco-Romance calques *ur pre začé* ‘11’, *doi začé* ‘20’, which by that time had been replaced in SIR by

7 The resistance of lower numerals against borrowing, observed in language after language, is probably rooted in the cognitively and genetically different substratum of numeric discrimination with small quantities (cf. e.g., the data on human infants and other animal species in Everett, 2017: 149–152; Vallortigara and Panciera, 2014: 52).

Slavic *jedennaist, dvaiset*,⁸ nowadays the only forms in use in both branches of IR.⁹

Apart from plain object-counting, any numerical expression that is even slightly conventionalized/culturalized tends to select Slavic numerals even more: thus, 'the Three Wise Men' is *tri kroki*, not **trej kroki*. The same goes for the quantification of time lapses and all time indications, where Romance numerals remain in use up to 4 only, as exemplified with 'hours' in (2):

- (2) NIR
- a. *na vedém na o ura* 'see you at 1 o'clock'
 - b. *na vedém na do/trei/patru ur* 'see you at 2/3/4 o'clock'
 - c. *na vedém na pet/*činč//šest/*švse//sedam/*švpte ur* 'see you at 5/6/7 o'clock'
(cp. Rom. *la orele cinci/șase/șapte* etc.)

In these phrases, also the preposition *na* is Slavic, as is the form of the noun meaning 'hour': *ura* NIR/*uræ* SIR: this is a reborrowing from Croatian (ultimately from Romance, cf. Kovačec, 1992: 161), but is used in (2b-c) in a form that does not match IR inflection (compare the IR plural *ure* 'hours') but rather corresponds to the Čakavian genitive plural (vs. Croatian *ura*) required in this context by Slavic syntax, which is taken on board in that case form (though, of course, with no case function). The same happens with any quantification expressed in terms of other temporal units (minutes, days, months, years etc.): e.g., *pet dân, šest dân, sedâm dân, ósâm dân* '5/6/7/8 days', *pet minut(i)* '5 minutes', *sedam miseți* '7 months', *deset let* '10 years' etc. Kovačec (1971: 218f.) shows that space and weight measures behave similarly, though here the Romance form for '5' can still be used: *pet kil/činč kile* '5 kilograms', *pet/činč métri* '5 metres'. He also reports the following dialogue with an informant:

Cînd, după ce am obținut de la ISV expresiile *pet dân, șest dân, sedâm dân, ósâm dân*, l-am întrebat dacă se poate spune și *činč zile* etc., răspunsul a fost: *betări re zice, ali âstez ți se re árde* 'bătrînii ar spune, însă astăzi ai fi luat în deridere' [When, after obtaining from ISV (= an informant from Žejane, born in 1902) the expressions *pet dân*, etc., I asked him whether one can also say *činč zile* etc., the answer was: 'old men would say so, but today you'd be mocked (for saying it)'].

8 Ascoli (1861: 75) actually writes "*dvaiste*", which might be a misprint, given that the Čakavian form for '20' is *dváiset* (compare Croatian *dvadeset*). NIR *dvajset ši ur/doj/trej* '21, 22, 23' were recorded in Ugo Pellis' fieldwork in 1926–1935 (cf. Dahmen and Kramer, 1988: 224).

9 Ascoli (1861: 75) also reports NIR *nuk* '9', not confirmed by any other sources and qualified as "obscure" by Pușcariu (1926: 153).

Thus, to the competence of a NIR informant interviewed when he was about 60, in 1961–1962, *činč zile* was a ludicrous archaism. Today, according to our informants, the Romance noun form *zile* (*zi* ‘day’ < Lat. DIEM) can be used in the phrase ‘five days’, to talk about *n* days qualifying them in terms of properties, but not in order to denote a time interval of *n* days (i.e., not as a time measure): one can say e.g., *činč zile fóst=av fine* ‘the five days were nice’, as opposed to *fóst=am ped dān in rika* ‘I spent (lit. was) five days in Rijeka’.

Sala (2013: 220), citing data from Kovačec (1966: 65f.; 1968: 99f.), Petrovici and Neiescu (1964: 19; also in 1965: 356), summarizes this situation as follows:

The co-presence of Romance and Croatian forms from ‘five’ to ‘eight’ is by no means a matter of ‘free variation’. Rather, the Croatian forms must be used in ‘lexical measure phrases’ (phrases expressing characteristic units of measurement, such as time, weight and distance); moreover, they must be combined with a Croatian noun, where one is available, showing Croatian noun morphology.

The same selectional restrictions on borrowed vs. native numerals now described with regard to measures hold true even when the word at issue (indicating e.g., a time lapse) is omitted, as in the exchange in (3a), where Slavic *pet* must be used even if *let* ‘years’ is gapped in B’s answer, to be compared with (3b), where in specifying the number of chickens, rather than years, **pet* is ungrammatical.

(3) a. A. *kəts* *vɲ* *aḷ* *fost* *la* *soldbót*
 how_much\M.PL year(M)\PL have.PRS.2SG been to soldier
 ‘how many years have you been in the army?’

B. *fóst=* *am* *samo* *doj/* *patru* (*vɲ*)// *samo*
 been= have.PRS.1SG only two.M/ four (year(M)\PL)// only
*pet/*činč* (*let*)
 five (years)
 ‘I’ve been there for only 2/4/5 years.’

b. A. *kət-e* *ɣačír* *vri*
 how_much-F.PL chicken(F)[PL] have.PRS.2SG
 ‘how many chickens have you got?’

B. *samo* *činč/*pet*
 only five
 ‘only five’

All this has been duly described in the literature on IR, and serves as background information to introduce the novel data on which our study focuses.¹⁰

5 Calqued and Borrowed Numerals with *Pluralia Tantum* Nouns in IR

The Romance languages, not unlike Latin and many other inflecting-fusional languages, have *pluralia tantum* nouns (henceforth PT). In Latin, as seen in (4), these nouns could be determined through the numeral ‘one’, in a context in which the morphosyntax of number (plural number, via agreement with the head noun) was in conflict with the semantics of the numeral, denoting *one* real-world entity:

- (4) *adeo* *ut* *un-a* *castr-a* *iam*
 to_the_point that one-NOM.N.PL camp(N)-NOM.PL already
fact-a *ex* *bin-is* *vid-ere-nt-ur*
 become-NOM.N.PL out_of two-ABL.PL seem-SBJ.IMPF-3PL-PASS
 ‘so that the two camps seemed to have become just a single one’ (CAES., *B.C.* 1.74)

At first glance, IR behaves like Latin in this respect (examples are from NIR; most of the phrases in (5) would be identical in SIR):

- (5) *ur-e* *brayeš-e/mudant-e/postol-e/ocvl-e/škr-e/novin-e/γrvbλ-i*
 (*or-e*)/*vil-e*
 one-F.PL trousers/underpants/shoe/spectacles/scissor/newspaper/
 rake/pitchfork(F)-PL
 ‘a pair of trousers/underpants/shoes/glasses/scissors//a newspaper/
 rake/pitchfork’

In (5), a series of PT nouns, all feminines, select the F.PL form of the numeral *ur-e* ‘one’.¹¹ Many of these nouns denote ‘objects made up of two like parts’ (Payne and Huddleston, 2002: 340; cited in Corbett, 2019: 54 n. 2) – e.g., *vil-e*

¹⁰ Though not in focus in the present paper, these facts are highly interesting per se, as they seem to represent a case of “parallel system borrowing” that could be added to those discussed e.g., by Kossmann (2010).

¹¹ Note that (4)–(5) show that, while occurrence of PT nouns in a singular indefinite context is rare cross-linguistically (cf. English **a pant(s)*, **a scissor(s)*), there are languages such as Latin or NIR which are exceptional in this respect, so that this cannot be regarded as a universal property of the grammar of PT nouns (*pace* Klockmann, 2017: 29).

‘pitchfork’ (a traditional pitchfork had two tines) –, which is a frequent case cross-linguistically for PT, though it need not be. In fact, *γρῦβλ-ι* ‘rake’ also denotes an object composed of a set of ordered parts, which are however more than two, and this semantic criterion hardly applies to *novine* ‘newspaper’ which parallels English *news* or its Hebrew equivalent *xadašót*, etymologically and morphologically. What crucially defines the nouns in (5) is that they have only plural morphology and invariably require plural agreement. Thus, they match, as to both the inflectional and syntactic criterion, Corbett’s (2019: 96) definition of PT as nouns that “have only the plural”.¹²

Since historically in Daco-Romance the Latin neuter plural has merged with the feminine plural (see Section 6, (27)), selection of the feminine plural quantifier in (5) could be regarded in principle as inherited from Latin. However, both the ecology of IR and the comparative (Daco-)Romance picture suggest a different explanation. Daco-Romance does not retain the numeral agreement pattern found in Latin (4), but rather replaces it with periphrastic classifiers, as exemplified with Romanian *pereche/perechi de* ‘pair/-s of’ in (6a):

- | | | | |
|-----|----|---|----------|
| (6) | a. | <i>o pereche/două perechi de ochelari</i> | Romanian |
| | | ‘a pair/two pairs of spectacles’ | |
| | b. | <i>un/doj p̄r de šk̄re</i> | SIR |
| | | ‘one pair/two pairs of scissors’ | |

This option is also available in IR (6b) while in Romanian it is compulsory, just as it is in Italian.

- | | | | |
|-----|----|------------------------------------|---------------------------|
| (7) | a. | <i>un paio di forbici/occhiali</i> | Italian |
| | | ‘a pair of scissors/spectacles’ | |
| | b. | <i>*une/-a forbici</i> | <i>*uni/-o occhiali</i> |
| | | one-F.PL/-F.SG scissors | one-M.PL/-M.SG spectacles |

In both Eastern Romance standard languages, plural forms of ‘one’ are never adnominal quantifiers but can only be indefinite pronouns/adjectives ((8a-b); an option available in IR as well, (8c)).

- | | | | |
|-----|----|--|---------|
| (8) | a. | <i>gli uni e gli altri / le une e le altre</i> | Italian |
| | | ‘the ones and the others(M/F)’ | |

¹² Corbett takes issue with definitions of PT nouns which refer to both form and meaning (e.g., “A noun which is plural in form but singular in meaning”, Trask, 1997: 172) and argues instead for a definition based on purely formal criteria (inflectional and syntactic).

- | | | |
|----|---|----------|
| b. | <i>unii ochelari</i> | Romanian |
| | ‘some (pairs of) spectacles’/*‘one pair of spectacles’ | |
| c. | <i>uri kār / ure vŋč</i> | NIR |
| | ‘some dogs/some cows’ | |
| d. | <i>unas gafas</i> | Spanish |
| | ‘some pairs of glasses’ or ‘(just) one pair of glasses’ | |

By contrast, Spanish has preserved the Latin option in (4), i.e., the pluralizability of ‘one’ with PT nouns, so that (8d), unlike its Italian and Romanian counterparts, is ambiguous. Plural agreement on the numeral quantifier ‘one’ with PT nouns is encountered occasionally in varieties which acquired it arguably via language contact. As the data in (9) indicate, Sissanese, a variety of Istrioto spoken in Sissano (South-eastern Istria, near Pula/Pola), is a case in point (see Giudici and Zanini, 2021).

- | | | | |
|-----|----------------------|-----------------|---|
| (9) | <i>un-e</i> | <i>fórfež-e</i> | Istrioto dialect of Sissano (Croatia) |
| | one-F.PL | scissors(F)-PL | |
| | ‘a pair of scissors’ | | (cp. Croatian <i>jedn-e škar-e</i> ‘id.’) |

These structural facts, along with the general attrition of IR under extreme contact, suggest that it is a priori more plausible to assume that the selection of the plural form of ‘one’ in IR with PT nouns such as those in (5) is a contact-induced phenomenon. The data in (10) display the Slavic model, of which (5) is most likely a replica.

- | | | | |
|------|---------------|---|----------|
| (10) | <i>jedn-e</i> | <i>hlač-e/gač-e/postol-e/naočal-e/</i> | Croatian |
| | | <i>škar-e/novin-e/grablj-e/vil-e</i> | |
| | one-F.PL | trousers/underpants/shoe/spectacles/scissor/newspaper/rake/pitchfork(F)-PL | |
| | | ‘a pair of trousers/underpants/shoes/glasses/scissors//a newspaper/rake/ pitchfork’ | |

Comparison with (5) reveals that most IR nouns in the latter example are loans from Croatian, including some words from the local Čakavian dialects such as *postol-e* (cp. standard Croatian *cipel-e*), and the two lists could be made even more alike if one considers that also *brayeš-e*, *mudant-e*, and *ocvl-e*, though ultimately of Latin descent, occur in Croatian dialects too and thus could be

13 Thanks to one anonymous reviewer for pointing this out to us.

Slavic borrowings just as well.¹³ The data in (11)–(12) display the Croatian paradigm from which the numeral form in (10) is picked, with one example for each gender/number value combination (all in the nominative). The examples in (12), from Leko (2009: 25) and Corbett (2019: 78f.), illustrate agreement with PT nouns.

- (11) a. *jedan* *tanjur* *je* *hladan* Croatian
 one[NOM.M.SG] plate(M)[NOM.SG] be.3SG cold[NOM.M.SG]
 ‘one plate is cold’
- b. *jedn-a* *kav-a* *je* *hladn-a*
 one-NOM.F.SG coffee(F)-NOM.SG be.3SG cold-NOM.F.SG
 ‘one coffee is cold’
- c. *jedn-o* *piv-o* *je* *hladn-o*
 one-NOM.N.SG beer(N)-NOM.SG be.3SG cold-NOM.N.SG
 ‘one beer is cold’
- (12) a. *jedn-i* *svat-ov-i* *su*
 one-NOM.M.PL wedding.procession(M)-PL-NOM.PL be.3PL
 stig-l-i
 arrive-PST-M.PL
 ‘one wedding procession arrived.’
- b. *jedn-e* *naočal-e* *su* *puk-l-e*
 one-NOM.F.PL spectacles(F)-NOM.PL be.3PL break-PST-F.PL
 ‘one pair of spectacles broke’
- c. *jedn-a* *kol-a* *su* *stig-l-a*
 one-NOM.N.PL carriage(N)-NOM.PL be.3PL arrived-PST-N.PL
 ‘one carriage arrived’

Thus, the IR F.PL form *ur-e* in (5) calques Croatian *jedn-e*. The table in (13) shows the complete paradigm which Kovačec (1971: 112) gives for the indefinite article.¹⁴

14 Most forms are homophonous with those of the numeral *ur* ‘one’, out of which they grammaticalized. Only in the nominative/accusative case, phonetic reduction is observed, which distinguishes M.SG *ən* and F.SG *o* in (13) from the numerals *ur/ura* in (1). The neuter form *uro* – whose *-o* ending is of Slavic origin, as discussed in Section 6 – is mostly used

(13)		M	F	N	The indefinite article in IR
SG	NOM/ACC	<i>ən</i>	<i>o</i>	<i>uro</i>	
	OBL	<i>urvé</i>	<i>urláé</i>	–	
PL	NOM/ACC	<i>uri</i>	<i>ure</i>	–	
	OBL	<i>uroré</i>	<i>uroré</i>	–	

While the occurrence of *ure* with PT nouns seen in (5) is observed in both NIR and SIR, the two branches of IR part ways as it comes to quantifying PT nouns with the numerals ‘two’, ‘three’ and ‘four’. The following examples are from Žejanski (NIR).¹⁵

- (14) a. *vl-ýt=am* *dvoje/*do/*doj* *škvr-e/γrɔbč-i/vil-e* NIR
 take-PTP=have.1SG two.X/two.F/two.M scissors/rake/pitchfork(F)-PL
 ‘I took two pairs of scissors/two pitchforks/rakes’
- b. *dæ=m* *troje/*trej* *novin-e/škvr-e*
 give.IMP=1SG three.X/three.M=F newspaper/scissors(F)-PL
 ‘give me three newspapers/pairs of scissors’
- c. *kumparót=am* *dvoje^{??}/do/*doj* *novin-e*
 buy-PTP=have.1SG two.X/two.F/two.M newspaper(F)-PL
 ‘I bought two newspapers’

All the feminine PT nouns in (14) require a special form of ‘2’ and ‘3’ which, as shown in (15), is distinct from the ones occurring with ordinary count nouns.

pronominally, but can marginally be used as an adnominal numeral quantifier or indefinite article as well.

- (i) *lá-a=ts* *ur-o* *srebro* SIR (Jesenovik)
 take-IMP.2SG=1SG one.N silver(N)
 ‘take one/a silver object’

15 The notation ‘two.X’, ‘three.X’ in the glosses will be explained in due course. In SIR, the ordinary feminine form is selected with such nouns, while **dvoje/*troje* are unacceptable: *do/*dvoje škvré* ‘two pairs of scissors’, *ste do/*dvoje ģornle* ‘these two newspapers’. Quantification of such nouns can also be realized periphrastically, as shown in (6b).

- (15) a. *trej//do/*doj* *pete/urec(Λ)i/bȃrba/ženske/măre/surăr/voč/
metle/fȃgate*
three//two.F/two.M heel/ear/beard/woman/hand/sister/cow/
broom/little girl(F)
- b. *trej//doj/*do* *oc(Λ)i/omir/dints/kăr/kúvete/piçore/kópure*
three//two.M/two.F eye/man/tooth/dog/elbow/foot/head(M)

While ‘three’ is invariable for gender in Daco-Romanian as well as in IR, ‘two’ inflects for gender in all Romanian dialects, as illustrated for IR in (15) (from now on, all examples will be from NIR whenever not otherwise specified).

Hence, the form of the numeral ‘two’ occurring in (14) is a third distinct form which adds to the two inherited ones. (The notation ‘two.x’ in the glosses means that its categorial status is still to be defined, an issue to which we will return in Section 8). While the form of ‘one’ occurring with PT nouns in (5) was arguably calqued on Croatian, *dvoje* and *troje* in (14) are borrowed, a fact that, to the best of our knowledge, has been mentioned so far, in the literature on IR, only in half a line by Puşcariu (1926: 156).

Substantive exprimând o parte sau o sumă de atâtea lucruri de acelaş fel găsim: *dvoje* (*biţvi*) < cr. *dvoje* şi (< ital.) *pâi* [...] ‘pereche’ [Nouns which refer to a part or a sum of several things of the same kind: *dvoje* (*biţvi*) ‘two pairs of socks’ and (from Italian) *pâi* ‘pair’]

Puşcariu’s wording and his quoting of a periphrastic classifier in the same context make clear that he is referring to the kind of quantification we are interested in. Curiously, his example is drawn from a SIR oral text collected in Šušnjevića in 1904 and printed in Puşcariu (1906: 180). Today, our SIR informants reject *dvoje* and *troje* categorically, in spite of using, of course, the homophonous collective numeral forms when they speak Croatian. This may perhaps indicate that the change whose results are evident in (14) was incipient at that time in SIR too, where however it did not become established.

In NIR, a borrowed form of the numeral occurs with the same nouns for ‘4’ as well, as seen in the series of examples in (16a), with feminine PT nouns, to be compared with those with plain count feminines in (16b).

- (16) a. *ure/ dvoje/ troje/ četvore/ *pétero/ *šestore novine/škȃre*
*o/ *do/ *trej/ (?)patru/ činč/ šȃve
‘one newspaper/pair of scissors//2/3/4/5/6 newspapers/pairs of scissors’

- b. *o mužbra* // *do/trej/patru/činč/špse mužér*
 ‘one wife’ // ‘two/three/four/five/six wives’

Since *čtvore* – also borrowed from Croatian – is uninflected, it will not detain us any further here. The examples in (17) show that our PT nouns consistently select feminine plural agreement on all agreement targets other than the numerals, illustrated here with demonstratives and qualifying adjectives.

- (17) a. *čol-e* *novin-e* *z betar-e* /**betar-i*
 DEM.DIST-F.PL newspaper(F)-PL are old-nom.F.PL /old-NOM.M.PL
 ‘those newspapers are old’
- b. *čvst-e* *postol-e* *s usk-e* /**usk-i*
 DEM.PROX-F.PL shoe(F)-PL are tight-NOM.F.PL /tight-NOM.M.PL
 ‘these shoes are tight’

The Slavic model is exemplified in (18), where the collective numeral adjectives (also termed “numerical adjectives”, see e.g., Lučić, 2015) are shown, which are selected with PT nouns of the three genders (data from Stefanović, 2014: 49; Lučić, 2015: 4f.; Kim, 2009: 114).

- (18) a. *ov-i* *dvoj-i/troj-i/četvor-i/peter-i* Croatian
 those-NOM.M.PL two/three/four/five-NOM.M.PL
svat-ov-i
 wedding.procession(M)-PL-NOM.PL
 ‘those 2/3/4/5 groups of wedding guests’
- b. *ov-e* *dvoj-e/troj-e/četvor-e/petor-e* *čarap-e*
 those-NOM.F.PL two/three/four/five-NOM.F.PL socks(F)-NOM.PL
 ‘those 2/3/4/5 pairs of socks’
- c. *ov-a* *dvoj-a/troj-a/četvor-a/petor-a* *kol-a*
 those-NOM.N.PL two/three/four/five-NOM.N.PL carriage(N)-NOM.PL
 ‘those 2/3/4/5 carriages’

In (18), for simplicity, only nominative forms are listed, since it is the F.PL nominative forms (*dvoje*, *troje*) that have been borrowed into NIR: the borrowing process probably started with whole NPs headed by PT nouns of Croatian origin such as e.g., *dvoje novine* ‘two newspapers’.

In the Slavic model system, those *e*-ending forms (18b) are homophonous with the non-agreeing collective numerals selecting genitive case on the noun they govern (Lučić, 2015: 5; Kim, 2009: 119).

- (19) a. *petero američk-ih studenat-a* Croatian
 five American-GEN.M.PL student(M)-GEN.PL
 ‘five American students’
- b. *troje telad-i*
 three calf(F)-GEN.PL
 ‘three calves’
- c. *dvoje djec-e*
 two child(N)-GEN.PL
 ‘two children’

This homophony is irrelevant structurally, because in terms of part of speech the model is (18b), not (19). However, it may have played a role enhancing the token frequency and hence the saliency of the forms *dvoje*, *troje*, *četvore*. All of the collective numerals in (18)–(19) are inherited from Common Slavic, were attested in Old Church Slavonic and are found in the modern languages, from Russian to Polish to BCS (cf. e.g., Brugmann, 1907: 22, 44f.; Kim’s, 2009 comparative syntactic study; or the further references in Corbett, 2019: 93). In modern spoken BCS they seem to be falling out of use at present according to the results of Stefanović’s (2011: 712) corpus study:

l’emploi normatif des adjectifs numériques, s’il se laisse observer ça et là, est peu vivant, relativement limité et tend à être remplacé par celui des numériques cardinaux, avec ou sans le lexème *par* « paire » [plus précisément, les adjectifs numériques (et les numériques collectifs) sont concurrencés par les cardinaux correspondants pour 2, 3, 4, tandis qu’à partir de 5, ce sont presque uniquement les cardinaux qui sont utilisés]. [‘the standard use of numeral adjectives, which one can observe at times, is not alive and well but rather limited and tends to be replaced by that of cardinal numerals, with or without the lexeme *par* ‘pair’ [more precisely, numeral adjectives (and collective numerals) are in competition with their cardinal counterparts for 2, 3, 4, whereas from 5 on, almost only cardinal numerals are used]’].

We are not aware of corpus-based studies on collective numerals in spoken Čakavian dialects, so we cannot speculate on their frequency of usage in the local contact varieties of NIR. Be that as it may, their existence in Čakavian – just as in standard Croatian (cf. e.g., Stevanović, 1989: 322 f.; Šipka, 2007: 121) –

is beyond doubt (*pace* Pranjković, 2000: 87): they are described by Ribarić (1940: 115) for Vodice, some 13 km WNW of Žejane and, as one anonymous reviewer kindly points out based on fieldnotes by Silvana Vranić, they occur in Mune Čakavian (3 km WNW of Žejane), the neighbour village's dialect used by Žejanski speakers (cf. Małecki, 1930: map. 4): e.g., *dvoje grablje/škare* 'two rakes/pairs of scissors', *dvoja kola* 'two cars', *četveroja vrata* 'four doors'. The same goes for the Čakavian variety of Orbanići, some 80 km to the WSW: *dvoji očenaši* 'two rosaries' (PT noun), *četvoreh postoli* 'of four pairs of shoes' (Kalsbeek, 1998: 178).

As said above, the first and only documentation of borrowed collective numerals in IR – the half line by Pušcariu (1926: 156) quoted above – goes back to the early 20th century, and the fact that it refers to SIR, where the change eventually aborted, may indicate that it was in the beginning at that time.

6 The Morphosyntactic System Into Which Collective Numerals Have been Borrowed

We now move on to discussing the impact that the borrowing of collective numeral forms from Croatian has had on the grammar of IR. In NIR, these forms have entered a grammatical system that distinguished two number values (singular vs. plural) and three gender values: masculine vs. feminine (inherited) vs. neuter (recently borrowed from Slavic into both IR branches), as exemplified with the paradigm of a class one adjective in (20a), compared with its Croatian counterpart in (20b) (Petrovici, 1967: 1525; Kovačec, 1971: 85).

(20) The paradigm of the adjective 'good'

a.	IR	SG	PL	b. Croatian	SG	PL
	M	<i>bur</i>	<i>bur-i</i>	M	<i>dobar</i>	<i>dobr-i</i>
	F	<i>bur-a</i> NIR/-æ	<i>bur-e</i> SIR	F	<i>dobr-a</i>	<i>dobr-e</i>
	N	<i>bur-o</i>	--	N	<i>dobr-o</i>	<i>dobr-a</i>

A number of studies have shown that the agreement marker *-o* occurring in IR class one adjectives (20a) was borrowed from Croatian, where it occurs in forms such as *dobr-o* in (20b). Once extracted, the affix applied to adjectives of the recipient language, including those from the inherited stock such as *bur-o*, resulting from *bur* (< Lat. BONUM) + *-o*. The introduction of this morph in IR occurred as new *o*-ending neuters such as *srebro* 'silver' entered the language without morphological adaptation, ousting the earlier strategy which produced adapted loans such as e.g., *okn-a* (NIR)/-æ (SIR) 'window(F)' < Sl. *okno* (Kovačec, 1998: 134; see what has been said in Section 4 on ancient loanwords

from Slavic into Daco-Romance, while commenting on borrowed *sută* ‘100’, and see Kovačec 1966: 67 on the replacement of this earlier strategy in IR with non-adapted borrowing of *o*-neuters).¹⁶

- (21) a. *złto=ȷ* *drvy-o*, *srebr-o* NIR
 gold(N)-SG=be.PRS.3SG expensive-N silver(N)-SG
nu=ȷ *drvg-o*
 NEG=be.PRS.3SG expensive-N
 ‘gold is expensive, silver is not’
- b. *tsăsta* *srebr-o=ȷ* SIR
 DEM.PROX:N silver(N)-SG=be.PRS.3SG
*vb-o/*vb*
 white-N/white[M.SG]
 ‘this silver is white’ (Šušnjevića, Pușcariu, 1906: 18, 39)

The *-o* ending seen in *drvy-o* and *vb-o*, selected by the Slavic borrowed nouns *srebro*, *złto* and the like, must have entered the language first in loan-adjectives such as *drvg-o*, to then spread to adjectives from the inherited stock such as *vbo* (from *vb* < Lat. ALBUM + *-o*), etc. Neuter *o*-forms of IR adjectives (including native ones, such as *buro* ‘good-N’, *groso* ‘big-N’) are reported as early as Pușcariu (1926: 15of.), quoting occurrences from the oral literature edited in Pușcariu (1906). Those occurrences, however, are invariably found in contexts in which no nominal controller is present, and can thus be interpreted as predicative adverbs (22a) or as instances of default agreement (22b).

- (22) a. *jel'-l' odgovorescu ke jel'-s siromăș si grumbo ke ăru*
 ‘they answer to him that they are orphans and are doing badly’
- b. *ța-i lăhco*
 ‘it’s easy’

Both uses are still observed today. In particular, the use of neuter agreement in default contexts, where there is no noun to trigger gender agreement, is obligatory.

16 Replacement of earlier adapted loans has been gradual. Thus, while Kovačec’s (1998: 225) dictionary only reports *zlăto* ‘gold(N)’ for Žejane, Kovačec (1963: 34) says that his Žejanski informants aged 50–70 used *złto=ȷ drvg-a* ‘gold(F) is expensive-F.SG’ and rejected as ungrammatical *złto=ȷ drvg-o* ‘gold(N) is expensive-N’, which was instead normally used by his younger informants (aged 12–17). We have recorded *złtăe* ‘gold(F)’, *złto=ȷ drvgăe* ‘gold(F).DEF is expensive’ from an informant from Šušnjevića born in 1954.

- (23) a. *vstez* *je* *yrumb-o* NIR
 today be.PRS.3SG ugly-N
 ‘the weather is ugly today’
- b. *tse=j* *de* *nov-o?* SIR
 what=be.PRS.3SG of new-N
 ‘what’s new?’

Comparing Pușcariu’s (1906) data with his own fieldnotes from the early 1960s, Kovačec (1963: 33–36; 1966: 67–70) showed that these *o*-forms had started to agree with nouns much later, and that the change was then still ongoing:

[C]ertains substantifs en *-o* empruntés au croate (emprunts probablement assez récents) ont commencé à s’accorder avec les formes adjectivales neutres en *-o* [‘some *o*-ending nouns borrowed from Croatian (probably relatively recent loans) have started to require, for agreement, *o*-ending adjective forms’] (KOVAČEC, 1966: 68).

Nowadays, as shown by ungrammatical **vb* in (21b), this agreement form must be categorically used with all and (almost) only the cited borrowed mass nouns. This is not just alliterative concord, given that borrowed mass nouns take the same *o*-agreement even if they do not end in *-o*, as long as they stay neuter, as shown for SIR in (24a) (the same Croatian loanword, on the contrary, has been recategorized as feminine in NIR because of its inflection class; see (24b)).

- (24) a. *uĭ-a=j* *drvȳ-o* SIR (Jesenovik)
 oil(N)-DEF.SG=be.PRS.3SG expensive-N/
**drvȳ-a/*drvȳ*
 expensive-F.SG/expensive[M.SG]
 ‘oil is expensive’
- b. *uĭ-a* *nu=j* *drvȳ-a/* NIR
 oil(F)-DEF.SG NEG=be.PRS.3SG expensive-F.SG/
**drvȳ/*drvȳ-o*
 expensive[M.SG]/expensive-N
 ‘oil is not expensive’

Conversely, neuter *o*-agreement with other non-neuter mass nouns – either native (such as *kvrne* ‘meat(F)’) or borrowed (such as *bronza* ‘bronze(F)’) – is generally judged ungrammatical.

- (25) a. *bronz-a* *nu=j* *ašá drvy-a/* NIR
 bronze(F)-SG NEG= be.PRS.3SG so expensive-F.SG/
 drvy*/drvy-o*
 expensive[M.SG]/expensive-N
 'bronze is not so expensive'
- b. *kvrn-a/svr-a* *=j* *bur-a/*
 meat(F)-SG/salt(F)-SG = be.PRS.3SG good-F.SG/
 bur*/bur-o*
 good[M.SG]/good-N
 'meat/salt is good'
- c. *unt-u/* *čest-a* *otsét* *je*
 butter(M)-DEF.SG/ DEM.PROX\M-SG vinegar(M)-SG be.PRS.3SG
bur/**bur-o*/**bur-a*
 good[M.SG]/-N/-F.SG
 'butter/this vinegar is good'

However, some exceptions – pointing to incipient semantic agreement – are reported by Kovačec (1966: 68):

Quelle est la pression du neutre croate, on le voit d'après le fait que deux substantifs *vir* e *lapte*, qui en istroroumain sont masculins, s'accordent quelquefois, sous l'influence des substantifs croates correspondants de genre neutre *vino* et *mlijeko*, 'faususement' avec les formes neutres des adjectifs. ['How strong the pressure of the Croatian neuter is, is seen from the fact that the two nouns *vir* 'wine' and *lapte* 'milk', which are masculine in IR, sometimes 'wrongly' take neuter agreement on adjectives under the influence of the corresponding Croatian neuter nouns *vino* and *mlijeko*].

The same vacillation is still observed in the competence and production of our informants:

- (26) *lɔpte-le* *nu=j* *drvy-o* */drvy* NIR
 milk(M)-DEF.SG NEG= be.PRS.3SG expensive-N /expensive[M.SG]
 'milk is not expensive'

As an output to the gradual spread now reviewed, the *o*-neuter has become established as a fully functional gender value. Since it is used in default contexts and with mass nouns which have been borrowed in the singular form

(with no plural), the neuter *o*-agreement in IR is number-defective, occurring in the singular only.

Since with reference to Daco-Romance the term neuter is usually employed to denote another distinct gender value, a word on the latter is in order here. Consider the (Daco-)Romanian gender system as illustrated in (27) with agreement on definite articles and a class one adjective (see Corbett, 1991: 151; Loporcaro, 2018: 92).

(27)	SINGULAR		PLURAL		Romanian
a.	<i>pantof-ul</i>	<i>e bun</i>	<i>pantofi-i</i>	<i>sunt bun-i</i>	
M	shoe(M)-DEF.M.SG	is good[M.SG]	shoe(M)-DEF.M.PL	are good-M.PL	
b.	<i>vin-ul</i>	<i>e bun</i>	<i>vinuri-le</i>	<i>sunt bun-e</i>	
N	wine(N)-DEF.M.SG	is good[M.SG]	wine(N)-DEF.F.PL	are good-F.PL	
c.	<i>băutur-a</i>	<i>e bun-ă</i>	<i>băuturi-le</i>	<i>sunt bun-e</i>	
F	drink(F)-DEF.F.SG	is good-F.SG	drink(F)-DEF.F.PL	are good-F.PL	
	'the shoe/wine/drink is good'				'the shoes/wines/drinks are good'

Nouns such as *vin* in (27b) are traditionally termed 'neuter' in Romanian descriptive grammar, which assumes that this is a third gender, distinct from both masculine and feminine. In Loporcaro (2018: 92–109), alternative two-gender analyses of Romanian are discussed and rejected, showing that a three-gender analysis is the only one in keeping with the following widely assumed definitions, which we assume here too.

- (28) a. "Genders are classes of nouns reflected in the behavior of associated words." (Hockett, 1958: 231; Corbett, 1991: 1)
- b. "We should [...] differentiate controller genders, the genders into which nouns are divided, from target genders, the genders which are marked on adjectives, verbs and so on." (Corbett, 1991: 151)

Under such definitions, the Romanian neuter, which is inherited from common Daco-Romance since it occurs in all of its four dialect branches (Petrovici, 1967: 1523), is a third controller gender, selecting agreement targets that are fully synthetic (with the masculine in the singular, with the feminine in the plural). As argued in Loporcaro (2018: 222), these syncretisms result from mergers. In other words, the Romanian neuter is inherited from Latin: only, it has turned from a target to a controller gender with alternating agreement.

Back to IR, while this language has acquired a new (mass) neuter via borrowing, by the time of Petrovici's (1962) study it had lost (NIR) or was in the process of losing (SIR) the inherited alternating neuter, which merges with the

masculine also in the plural. As a result, nouns like those in (29), which used to select alternating agreement (and still do in Daco-Romanian, (27b)), now select plural masculine agreement in NIR and consequently have been reassigned to the masculine.

(29)	<i>češc-i/*čvšt-e</i> DEM.PROX\M-M.PL/DEM.PROX\F-F.PL <i>kóp-ure</i> head(M)-PL ‘these two elbows/heads’	<i>doj/*do</i> two:M/F	NIR <i>kúvet-e/</i> elbow(M)-PL/
------	--	---------------------------	--

Thus, IR shows that contact-induced pressure may result not only in the simplification of grammar, even if the latter is most often the case cross-linguistically: “language contact, especially when extensive L2 learning is involved, is a main source of complexity reduction (grammar simplification)” (Karlsson et al., 2008: viii; see also Arkadiev and Gardani, 2020). On the contrary, the rise of the *o*-neuter in IR is a case of contact-induced complexification, rather than simplification, of a language’s grammar (on a par with the others discussed, with reference to gender, in Loporcaro, 2018: 51f.).

7 Intermezzo: the Slavic neuter and the Double Complexification of the NIR Gender System

In Section 6, we have seen that the *o*-neuter had already been borrowed by the time grammatical descriptions of IR became available and acquired its syntactic function of conveying agreement with neuter *o*-loans by the mid-20th century. A still later development was first reported as interference occurring in the performance of some (mostly rather attrited) Žejanski speakers by Kovačec (1963: 35).

Nous n’avons pas rencontré de pluriels de substantifs neutres, sauf dans deux cas douteux. Une jeune fille de 21 ans, qui a vécu assez longtemps à Rijeka où elle faisait ses études, en traduisant un texte croate a employé *ašava pitaña* ‘de telles questions’ comme pluriel. L’autre exemple, si l’on applique des critères croates à l’analyse, ne pourrait que confirmer indirectement l’existence d’un embryon du pluriel neutre. Pour ‘les enfants étudient’ nous avons noté à Žejane *diša se-nveštavę* et *diša se-nveštavés* (pl. neutre ?), mais il pourrait s’agir ici seulement d’un calque du pluriel

croate dans la forme verbale [...], le substantif étant pris comme un collective de genre féminin (ce qui se rencontre quelquefois dans les dialectes čakaviens environnants: *dica se uči* à côté de *dica se uču, učiju* [...]) [‘I did not come across any plurals of neuter nouns, except for two dubious cases. A 21-year-old girl, who lived for a relatively long time in Rijeka where she was studying, used, in translating a Croatian text, *ašáva pitańa* ‘such questions’ as a plural. The other example, if analysed by Croatian criteria, could not but confirm indirectly the existence of an embryo of the neuter plural. For ‘children study’ I have recorded in Žejane *diťa se-nveřavé* and *diťa se-nveřavés* (neuter plural?), though it could be nothing else than a calque of the Croatian plural in the verb form [...], whereby the nouns could be taken as a feminine-gender collective (which is met with at times in the neighbouring Čakavian dialects: *dica se uči* ‘children learn:SG’ alongside *dica se uču, učiju* ‘children learn/are learning:PL’ [...])’]

Kovačec’s (1998: 69) dictionary follows this latter interpretation (singular “collective” noun with semantic plural agreement) in specifying the entry *di:tsa* ‘children’, as follows: “**diťa** ž (zbirno)” [‘**diťa** f(eminine) (collective)’]. The same grammatical specification is given for *γospoda* ‘(wealthy) gentlemen’, reported in Kovačec (1998: 85) alongside a separate entry for *γospodvr* and *γospodín* ‘seigneur’, while *di:tsa* lacks a similar M.SG counterpart altogether.

The agreement pattern described as occasional interference and/or semantic agreement by Kovačec (1963: 35) has now become established in NIR, where the cited nouns – plus *vlastela* ‘noblemen’, not registered in Kovačec’s (1998) dictionary – select unambiguously a type of syntactic agreement that was not observed in IR prior to borrowing. In today’s NIR, there is little doubt that those three lexemes must be regarded as PT nouns, for they obligatorily select plural verb agreement, as the data in (32) and (33) show. This is in keeping with their origin, as they are all homophonous with the model Croatian forms, among which *dica* is the local Čakavian dialect variant (vs. standard Croatian *djeca*) with *i* < Proto-Slavic **ě* found in the dialects of the Mune area (cf. Mañeckí, 1930: map. 4). In the source language, these are plurals from non-defective paradigms (*djeca*) or can occur with either plural or singular agreement (*γospoda*).

In the whole of BCS, the noun for ‘children’ presents an intriguing and much-discussed situation: while it serves as a plural to *dijete/dete* ‘child(N)’ and has plural morphosyntax (i.e. agreement), morphologically it inflects like the singular of feminine nouns ending in *-a* (such as e.g., *žena* ‘woman’; cf.

Corbett, 1983: 76–81; 2000: 187f.; 2007: 39; Despić, 2017): this is seen in (19c) above, where *djec-e* is genitive plural morphosyntactically but has an *-e* ending which corresponds morphologically to the genitive singular of the feminine *a*-class: compare *žene* vs *ženā* ‘woman’ (GEN.SG vs. GEN.PL). By contrast, *gospoda* selects either singular or plural agreement, with a semantic difference (‘gentry’ vs. ‘lords’) discussed in Stankiewicz (1983: 157):

- (30) a. *gospod-a* *se* *nije* *meša-l-a*
 gentry(F)-NOM.SG REFL NEG:be.PRS.3SG mingle-PST-F.SG
s *narodom*
 with people(M)-INSTR.SG
 ‘the gentry did not mingle with the people’
- b. *gospod-a* *se* *jesu*
 gentry(N)-NOM.PL REFL be.PRS.3PL
zavadi-l-a
 quarrel-PST-N.PL
 ‘the lords had a falling out’

The three nouns behave differently in the two IR branches. Our SIR informants do not accept *dirtsā* and *vlastela* as possible IR words but do use *gospoda* – in exactly the way reported in Kovačec (1998: 85) only for NIR – as a singular with collective meaning. This can be predicated of a plural NP, as shown in (31a), but when employed as a subject never takes plural verb agreement (31b).

- (31) a. *ječ=əz* *gospod-a* NIR
 3M.PL=be.PRS.3PL gentlemen
 ‘they are (wealthy) gentlemen’/‘they are gentry’
- b. *ts-a* *gospod-a=j/*s* SIR
 DEM.DIST-F.SG gentlemen(F)-SG=be.PRS.3SG/.3PL
čvro *bogpt-a*
 very rich-F.SG
 ‘those gentlemen are very rich’

Thus, in borrowing this lexeme, SIR selected one of the two options Croatian offered, viz. (31a). By contrast, NIR took the other option, (31b), as all the three above mentioned *a*-plurals, including *gospoda*, select plural agreement on verb forms, as in the model language, as shown in (32), while at the same time selecting an *a*-ending on other agreement targets which – as first remarked in Loporcaro (2018: 294f. n. 6) – is never found elsewhere in the language, where

the inherited paradigm of plural agreement targets in the relevant inflectional class(es), as seen in (20a), maximally features the binary contrast *bur-i* ‘good-M.PL’ vs. *bur-e* F.PL.¹⁷

- (32) a. *č-a/*čeŕ-i* *gospod-a=s/*j* NIR
 DEM.DIST-N₂\-M.PL gentlemen(N₂)-PL=be.PRS.3PL/.3SG
jvko *bogpt-a/bogóts*
 very rich-N₂/rich\M.PL
 ‘those gentlemen are very rich’
- b. *č-a* *gospod-a=z/*j*
 DEM.DIST-N₂ gentlemen(N₂)-PL=be.PRS.3PL/.3SG
*nægr-a/negr-i/*nægr-e*
 black-N₂/black-PL/black-F.PL
 ‘those gentlemen are black’
- c. *č-a* *pərv-a* *dits-a* *mi=z*
 DEM.DIST-N₂ first-N₂ child(N₂)-PL 1SG.DAT=be.PRS.3PL
mai *bur-a²bur-i*
 more good-N₂/good-M.PL
 ‘I prefer those former children’
- d. *bogpt-a vlastel-a* *ən* *mostar av* *fost grvs-a*
 rich-N₂ noblemen(N₂)-PL in Mostar have.PRS.3PL been fat-N₂
 ‘the rich noblemen in Mostar were fat’
- e. *čvst-a* *dits-a=z* *mik-a/*mič*
 DEM.PROX-N₂ child(N₂)-PL=be.PRS.3PL small-N₂/small.PL
 ‘these children are small’

This *a*-ending is exemplified for qualifying adjectives and demonstratives in (32), to which relative pronouns are added in (33).

- (33) a. *čvst-a* *dits-a* *kar-a/kar-ŕi* *av* *verít*
 DEM.PROX-N₂ child(N₂)-PL REL-N₂/REL-M.PL have.PRS.3PL come
ən *sélište av* *fost* *visóč*
 in village have.PRS.3PL been tall.PL
 ‘these children that came to the village are tall’
- b. *dits-a* *kvr-a/*kvr-ŕi* *am* *vezút*
 child(N₂)-PL REL-N₂/REL-M.PL have.PRS.1SG seen
 ‘the children that I saw’

17 Contrary to their Daco-Romanian counterparts, NIR *mik* ‘small’ and *negru* ‘black’ inflect differently, as the plural forms *mič* and *negri* are used for both masculine and feminine agreement (cf. Kovačec, 1998: 116, 126). However, for the latter adjective, while our informants indeed use *negri* for both genders, they also have a dedicated F.PL form *nægr-e*, which is

With these nouns, *a*-agreement is always acceptable while feminine plural *-e* never is (see (32b)). As far as masculine plural agreement is concerned, this is sometimes deemed fully acceptable (32b) and (33a), sometimes regarded as dubious (32c), sometimes excluded (33b). An Agreement-Hierarchy effect (Corbett, 1979), whereby M.PL agreement is the more acceptable the further away from the NP-internal attributive position, seems to be suggested by (33a), but (33b) (a judgement by the same informant) is not in line with this speculation. The crucial point for our reasoning is that verb agreement is always plural. Were this not the case, NIR would be like SIR (see (31)) or Sursilvan (see §8.1). But since these are undoubtedly plural nouns, and they select an agreement morph which never occurs with M.PL and F.PL nouns, these nouns must be specified for a distinct gender value, which is notated N(EUTER)₂ in the glosses, to distinguish it from the *o*-neuter seen in (20)–(23). This means that, taking the data in (32)–(33) into account, one needs to further complexify the gender system of NIR with respect to what available grammars have said so far. Our analysis is provisionally schematized in (34a).

(34) Gender/number agreement in IR

a. NIR	SG	PL	b. SIR	SG	PL
M	<i>bur</i>	<i>bur-i</i>	M	<i>bur</i>	<i>bur-i</i>
F	<i>bur-a</i>	<i>bur-e</i>	F	<i>bur-æ</i>	<i>bur-e</i>
N ₁ (MN)	<i>bur-o</i>	–	N (MN)	<i>bur-o</i>	–
N ₂ (CN)	–	<i>bur-a</i>			

While the *o*-neuter₁ – which could be alternatively labelled M(ASS) N(EUTER) – is syntactically productive, as seen from the fact that it has taken on the default function, the *a*-neuter₂ – or, alternatively, C(COLLECTIVE) N(EUTER) – is not: rather, with just three borrowed nouns assigned to it, it is a vanishingly small gender value which, however, must be recognized as such. In particular, by Corbett's (2012: 84) criteria, one cannot call it an inqorate gender, since inqorate gender values are those “which comprise a small number of nouns, and whose agreements can be readily specified as an unusual combination of forms available for agreement with nouns with the normal gender values”. The relevant cases reviewed in Corbett (1991: 170–175) are all instances of controller genders with no dedicated agreement targets, which – if the numbers are very small (one or two lexemes) – may be treated alternatively as lexical exceptions. Neither of these alternatives is available for the NIR neuter₂, since its agreement marker *-a* is a dedicated one, as no other word in the language selects it

ungrammatical with *a*-plurals as seen in (32b) but can occur e.g., in *čale do fete=z negr-i/nægr-e* ‘those two girls(F) are black’.

in the plural. Thus, it matches the requirement put by Corbett (2012: 84, fn. 12): “If such nouns have their own unique agreement forms, rather than taking a combination of forms which are otherwise available, the agreement class must be recognized as a gender value, even if few nouns are involved.”

As to the origin of this *a*-ending, it is clear – as Kovačec (1963: 35) remarks – that it is ultimately due to Croatian influence: if the two developments in SIR and NIR are independent from each other, in the latter this *-a* may have been extracted from phrases such as *bogata gospoda* ‘rich gentlemen/lords’, *draga dica* ‘dear children’ (for the mechanisms of direct vs. analogical borrowing of inflectional morphemes, cf. Gusmani, 1979; Gardani, 2008; 2012; 2018; 2020; Seifart, 2015).

The *a*-collectives assigned to the neuter₂ all share the property of not being determinable through numerical quantifiers (35a), a situation which is encountered sometimes, across languages, with collective nouns (cf. Loporcaro, 2018: 73f., for discussion of a parallel from Romansh; as (35b) shows, other quantifiers are not barred, and they regularly agree in *-a*).¹⁸

- (35) a. **ura* // **do*/ **doj*/ **dvoje*/ **dvoja* *di:ts-a*
 one:*N₂// two.F/ two.M/ two:FC/ two:*N₂ child(N₂)-PL
 (intended ‘one child/one group of children//two children’)
- b. *núškvr-a* *di:ts-a*
 some-N₂ child(N₂)-PL
 ‘some children’

For SIR, the schema in (34b) is not complete, since it displays the three target genders but omits the inherited alternating neuter (= AN in the gloss in (36a)), which has persisted longer in this IR branch (see discussion on (29)).

- (36) a. %*do* *kúvat-e/žóžet-e/lúpure* SIR
 two.F elbow(AN)-PL/finger(AN)-PL/wolf(AN)-PL
- b. *doj* *kuvats/ždžets/lup*
 two.M elbow(M)\PL/finger(M)\PL/wolf(M)
 ‘two elbows/fingers/wolves’

18 Among the numeral forms in (35a), *do*, *doj*, *dvoje* are used in other contexts in NIR, and therefore exist, as the reader by now knows, whereas **ura* and **dvoja*, to the best of our knowledge, do not (which is signalled by the asterisk included in the glosses “one:*N₂” and “two:*N₂”). The latter forms have been built with the intention of exploring the theoretical possibility for speakers to create forms with the appropriate inflection for that feature-value combination, to be used with *a*-collectives, and, for *dvoja*, based on the homophonous NOM.N.PL form of the Croatian collective numeral (see (18c)).

As a matter of fact, at least one of our informants from Šušnjevića still variably allows F.PL, alongside innovative M.PL agreement (36b) – nowadays prevailing – with original Daco-Romance neuter nouns such as *kúvat* ‘elbow’ and *žbžet* ‘finger’, an option that, before it started to beat a retreat, had been extended even to original masculines such as *lup* ‘wolf’.

To sum up, contact-induced complexification of the gender system seems to be on the rise in IR. The changes that led to the emergence of the two neuters in NIR are clearly contact-induced. Interestingly, two values of one and the same gender in the source language (neuter singular and neuter plural) have been copied at different times in the recipient language, not as part of one and the same paradigm but rather as two distinct defective gender values.

8 Borrowed Numerals with *PT* and the Further Complexification of the Gender System

Back to numerals, let us now consider the impact of the borrowing of Croatian *dvoje*, *troje* on the morphosyntactic system described in Sections 6–7. As shown in (37), this borrowing has turned a formerly binary option in the agreeing forms of the numeral ‘two’ into a three-way one, whereas all other agreement targets – exemplified in (37) with the demonstrative – only contrast two forms in the plural, in the paradigms usually given by grammars (see the demonstrative paradigm in Kovačec, 1971: 109, to which the NIR N_2 is added in (38)):

- (37) a. *češč-i* *doj/*do/*dvoje* *omir/dints/kər/kúvete* NIR
 this-M.PL two.M/two.F/two.X man/tooth/dog/elbow(M):PL
 ‘these two men/teeth/dogs/elbows’
- b. *čvst-e* *do/*doj* *surár/vvč/metle/fčate*
 this-F.PL two.F/two.M sister/cow/broom/little_girl(F):PL
 ‘these two sisters/cows/brooms/little girls’
- c. *čvst-e* *dvoje/*do/*doj* *novin-e/postol-e/vil-e/škor-e*
 this-F.PL two.X/two.F/two.M newspaper/shoe/pitchfork/scissors(F)-PL
 ‘these two newspapers/shoes/pitchforks/pairs of scissors’

(38) The proximal demonstrative in NIR:

		M	F	N ₁	N ₂
SG	NOM/ACC	čest-a	čvst-a	čest-a	
	OBL	čest-váé	česc-áé	–	
PL	NOM/ACC	češc-i	čvst-e		čvst-a
	OBL	čest-oráé	čest-oráé		–

The forms *dvoje* and *troje* are by now well integrated in the recipient system, so much so that, having been stripped away from the Croatian inflectional paradigm and having thus lost all the case/gender/number endings other than *-e*, they have developed oblique case forms by analogy with the nominal oblique endings of IR (compare e.g., *hartá novinelor je raskinita* ‘the paper of the journal is torn’):

- (39) a. *várx-ur-le* *dvoje(r)lor* *škvr-e*
 tip(M)-PL-DEF.PL two.X:OBL.PL scissors(F)-PL
 ‘the tips of two pairs of scissors’
- b. *roba* *trojerlor* *mudant-e*
 cloth(F)-SG three.X:OBL.PL underpant(F)-PL
 ‘the cloth of three pairs of underpants’

Synthetic oblique endings for nouns and pronouns were lost altogether in SIR and only preserved in NIR (Petrovici and Neiescu, 1965: 360). Among numerals, this is the case only in *ur* ‘one’, as shown in (13), while the others, including *do/doj* ‘two’, form the oblique case periphrastically preposing the case marker *a*: *a do/doj* ‘two.OBL.F/M’ (Kovačec, 1971: 117). Against this background, the morphological integration of *dvoje* and *troje* shown in (39) appears all the more remarkable.

In what follows, we are going to explore the idea that also the borrowing of the numerals *dvoje* and *troje*, not unlike that of the *o*- and *a*-neuter agreement markers considered in Sections 7–8, may have increased the complexity of the recipient morphosyntactic system. We argue namely that this borrowing resulted in introducing gender overdifferentiation into the paradigms of the two agreement targets at issue. In other words, we propose that the three-way contrast seen in (37) has to be treated as one of (sub)gender.

8.1 Comparative Evidence: Gender Overdifferentiation on Lower Numerals in Romance

The typological presupposition of what we have been saying so far is that “[c]ardinal numbers sometimes show agreement; typically, this is restricted to lower numerals” (Corbett, 2006: 42). Lower numerals, cross-linguistically,

have also not rarely been reported to host gender overdifferentiation, defined as follows:

for targets to be considered overdifferentiated, a specific gender agreement distinction must be restricted to a particular word-class, and even within this word-class it must be restricted to certain lexical items. (CORBETT, 1991: 169)

Corbett (1991: 168f.) cites examples of overdifferentiation on the numerals ‘two’, ‘three’ and ‘four’ in Kolami-Naiki and Parji-Ollari, two central Dravidian languages in which only those numerals display dedicated agreement forms for female human nouns, in addition to those occurring on all other agreement targets, which contrast only male human vs. other. In Romance, a comparable state of affairs is observed in Romansh, as shown with examples from Sursilvan and Engadinian in (40) (data from Candinas, 1982: 110f.; Spescha, 1989: 312f., on Sursilvan; Ganzoni, 1977: 56f., on Upper, and Ganzoni, 1983: 56f. on Lower Engadinian):

(40) a. Srs.	b. Eng.	gloss
<i>du-s/trei-s mattatschs</i>	<i>du-os/trai-s mats</i>	‘two-M/three-M boys(M)’
<i>du-as/trei-s mattatschas</i>	<i>du-os/trai-s mattas</i>	‘two-F/three-F girls(F)’
<i>du-a/trai-a pèra</i>	<i>du-a/trei pèra</i>	‘two-N/three:N pairs(?)’

In addition to masculine and feminine, generally contrasted on plural agreement targets from all relevant classes, the numerals ‘two’ and ‘three’ feature a distinct form ending in *-a* – a diachronic successor of the Latin neuter plural agreement morph *-a* – which nowadays only occurs within complex numerals such as Eng. *duatchient* ‘200’, *traiamilli* ‘3000’ and the periphrastic quantifiers in (40).¹⁹ However, until not long ago these forms could modify *a*-collectives like *bratsch-a* ‘arms(F)-SG’ even in their literal meaning.

- (41) *Tgi che ha duas combas e dua bratscha duei gie buca selubir da simular e mulestar il miedi* (Candinas, 2009: 91)
 ‘that who has two legs and two arms should not dare to simulate and disturb the doctor’

¹⁹ While *duo* was the Classical Latin neuter form, an analogical variant *dua*, with the nominative/accusative ending reshaped on the model of nominal inflection, is also attested: see *ThLL*, 5(1): 2241f.

The author of the novel from which the passage is drawn, Theo Candinas, was born in Surrein-Sumvitg, Surselva, in 1929; for younger speakers, *dua bratscha*, if at all acceptable, can only denote a measure, meaning ‘two ells’ (see Kämpf, 2015).

Exactly the same overdifferentiation on the numerals ‘two’ and ‘three’ occurred in medieval Northern Italo-Romance (in Veneto, Lombardy, Emilia and Liguria: see Loporcaro and Tomasin, 2016) where these were the only plural agreement targets to feature a three-way gender distinction:

- (42) Gender overdifferentiation on ‘two’ and ‘three’ in medieval Northern Italo-Romance

	M	F	N	
PL	<i>dui, tri</i>	<i>doe, tre(i)</i>	<i>doa, trea</i>	‘two, three’
	<i>li</i>	<i>le</i>	–	DEF ART (= all other AGR targets)

Here too, the *a*-forms could not modify normal count nouns but were restricted to use within periphrastic quantifiers (‘two/three pairs of *x*’) and complex numerals: (page numbers are given in brackets):

- (43) 14th century Venetian (*Tristano corsiniano*, edn. Tagliani, 2011)
- M *li dui fradelli* (73) ‘the two:M brothers(M)’; *li altri dui compagnon* (89) ‘the other two:M companions(M)’; *delli dui servi* (99) ‘of the two:M servants(M)’;
- F *tute doe le palme* (108) ‘both [lit. ‘all:F two:F’] palms(F.SG)’; *a doe mane* (108) ‘with both [lit. ‘two:F’] hands(F)’; *doe çornade* (144) ‘two:F days(F)’;
- N *doa para d’arme* (124) ‘two:N pairs of arms/armours’; *plu de doa milia* (145) ‘more than two:N thousand’; *doa tanti* (127) ‘twice [lit. ‘two:N’] as many’.

Of course, the data discussed in (40)–(43) differ from those from NIR in several respects. On the one hand, diachronically, overdifferentiated forms are inherited in Romansh, as they were in medieval Northern Italo-Romance, being a leftover of the Latin three-gender system which elsewhere shrunk to a binary contrast; in NIR, on the contrary, they arose from language contact. Synchronically, on the other hand, those seen in (40)–(43) are plain three-way contrasts, whereas in NIR the situation is, also in this respect, more complex.

However, there is also a striking similarity. While the data in (41) still bear witness to the original plurality of the *a*-noun forms selecting *dua* and *trei/traia*, such noun forms in modern Romansh belong to number-defective paradigms with a form/meaning mismatch: Sursilvan *bratscha* denotes two entities but is morphosyntactically singular, a mirror image with respect to NIR PT nouns selecting *dvoje/troje* such as *novine* ‘newspaper’.

8.2 Contact-Induced Gender Overdifferentiation for Lower Numerals in NIR

The scheme in (44) displays the usual situation for gender/number marking on (non-overdifferentiated) agreement targets exemplified with the paradigm of *ur* ‘one, some’:

(44)

	M	F	N ₁	N ₂	
SG	<i>ur</i>	<i>ur-a</i>	<i>ur-o</i>		‘one’
PL	<i>ur-i</i>	<i>ur-e</i>		–	‘one, some’

In addition to the contrasts seen in (44) – two number and three gender values (no plural **ur-a* occurs, as the neuter₂ never occurs with numeral quantifiers, see (35a) and fn. 18) – the schema in (45) adds complexity in the form of a layering in the feminine (here, also the N₁ does not occur, since the agreement targets at issue are plural while the N₁ only occurs in the singular):

(45)

	M		F		N ₁	N ₂	
			collective				
PL	<i>doj</i>	<i>do</i>	<i>dvoje</i>			–	‘two’
	<i>trej</i>		<i>troje</i>			–	‘three’

We know independently (see (17) and (37c)) that PT nouns which select *dvoje/troje* are feminine and plural, and that they share this feature specification with ordinary count feminines that select inherited *do* ‘two.F’ instead (37b). Thus, they all share the same gender/number specification, so that our hypothesis is that overdifferentiation in lower numerals signals what has come to be a subgender contrast in NIR. In (45), the subgender signalled by selection of *dvoje*, *troje* is labelled ‘collective’ in a merely conventional way: while this alludes to the origin of the borrowed agreeing numerals, it does not imply retention of the original semantics of collective numerals, a point to be dealt with in Section 8.3.

Synchronically, we argue that borrowed *dvoje* and *troje* are now distinct word forms in one and the same paradigm together with the inherited forms of the numerals ‘two’ and ‘three’ (the non-greyed-out cells in (45)): in other words, though differing in origin, native *doj/do* and borrowed *dvoje* have become part of one and the same numeral lexeme, and the same goes for native *trej* and borrowed *troje*.

At this point, a series of questions arise, whose discussion will require considering additional comparative evidence from Romance and beyond: a) firstly, and crucially, the question whether this idea is on the right track, considering that no such morphosyntactic analysis has been proposed yet, to the best of our knowledge, for the many languages in which PT nouns select special numerals;

b) secondly, the issue whether, in case overdifferentiation is assumed, this is best analysed in terms of (sub)gender, or whether it should rather be treated in terms of some other morphosyntactic feature; c) thirdly and finally, whether – assuming the (sub)gender analysis is correct – the gender-asymmetry seen in (45) is justified, or whether such overdifferentiation should rather be assumed also for the masculine. We will start by discussing the last issue in Section 8.3, since the data introduced there will pave the way for addressing the fundamental issue (a) in Section 8.4, where quantification with Latin PT nouns will be drafted in as a useful comparison. Finally, in Section 8.5 we will show that the analysis in terms of (sub)gender is preferable over conceivable alternatives appealing to other morphosyntactic features.

8.3 *Lack of Overdifferentiation in the Masculine and the Semantics of Dvoji and Dvoje*

Kovačec's (1998) dictionary contains a handful of masculine nouns, whose lexical entries are given in the plural and may consequently stand as candidates for PT status. These all reported in (46):

- (46) Masculine plural entries in Kovačec's (1998) dictionary
 [legend: B = Brdo (SIR), J = SIR (all villages), S = Šušnjevića (SIR),
 Ž = Žejane (NIR), It[alian], Ven[etan]:
armi (J) 'weapons' (It. *armi*), *bizi* (S) 'peas' (Ven. *bizi*), *boše -le* (Ž)
 'testicles',²⁰ *cârmel' -i* (J Ž) 'pebbles of sleep dust' (Cr. *krmelji*), *cvadri*
 (B) 'sacred images' (It. *quadri*)

As is readily apparent, most of them are not used in Žejanski but only occur in SIR, so that only NIR *boše* and *cârmel'* are potentially relevant to our question. We have tested them, asking our informants whether they could be quantified with M.PL *dvoji* (see the possible Croatian source *dvoji* in (18a)), with the following results:

- (47) *doj/*dvoji boše*
 'two testicles'
- (48) a. *doj kârmel'*
 'two pebbles of sleep dust'
 b. [%]*dvoji kârmel'*
 'two pebbles of sleep dust'

²⁰ Note that the lexeme *boše, -le* (Ž) has a plural entry in Kovačec's (1998: 40) dictionary. However, the author also cites sg. *an boš* 'a testicle'.

Most speakers reject the sentences with *dvoji* in (47) and (48b) as ungrammatical. For two of our informants, however, *dvoji kərmeɫ* (48b) is acceptable, although only if the objects belong to two different sets of pebbles of sleep dust, e.g., one/two from one eye, one/two from the other. The remaining speakers reject it outright. The crucial point for us is the fact that *dvoji* with these nouns is not selected categorically as the only grammatical form of the numeral ‘two’, contrary to what is observed with the feminine PT nouns in (14) and (16), nor do any other masculine PT nouns seem to exist for which this would be the case. This guarantees that (45) is correct in not positing any subgender contrasts for the masculine: in other words, the contrast between F *do* and *dvoje* in (15) vs. (14) is relevant to the morphosyntax, while the difference between M *doj* and *dvoji* (for the NIR speakers who deem the latter form grammatical, in (48b)) never is.

Indeed, also the *do* ≠ *dvoje* contrast in the feminine may convey, with non-PT nouns, a purely semantic contrast not relevant to the morphosyntax like the one seen in (49). In fact, while feminine PT nouns select the numerals *dvoje*, *troje* categorically, the latter are not restricted to quantification of PT nouns, but can also quantify countable plurals, exemplified with *ženska* ‘woman’ and *šalitsa* ‘cup’ in (49b) and (50b):

- (49) a. *trej ženske* ‘three women’ (unmarked)
 b. *troje ženske* ‘three women’ (= three different [sets of] women)
- (50) a. *vezút=vm do/trej šalitse* ‘I’ve seen two/three cups’
 (unmarked)
 b. *vezút=vm dvoje/troje šalitse* ‘I’ve seen two/three cups’
 (of different kinds)

When this happens, these expressions, contrary to those with cardinal numerals in (49a), (50a), indicate that what is being referred to is either two/three sets (for some speakers) or two/three items only if picked from distinctly different sets (for others):²¹ for speakers of the former group, *dvoje šalitse* means ‘two sets of cups’, independently of the number of items in each group. The same usage of collective numerals is observed with masculine count nouns too:

21 For instance, (49b) may indicate – for speakers of the former group – that the three women at issue are instances of different types e.g., in that they come from the set of red-haired, black-haired, and blond women.

- (51) a. *doj kǔl* 'two horses' (unmarked)
 b. *dvoji kǔl* 'two different [groups of] horses'

- (52) a. *trej kǎr* 'three dogs' (unmarked)
 b. *troji kǎr* 'three different [groups of] dogs'

In the light of this, (48b) does not seem to instance the kind of morphosyntactically obligatory use the borrowed numerals have been put to in NIR, described in (14), (16) and (37c). Rather, it seems to be interpretable as a manifestation of the same optional collective use found in the source language: the possibility to convey such 'group' meaning is part of the rich semantics of Slavic collective numerals (see Leko, 2009: 76–81 and Stefanović, 2011 for BCS).

8.4 *A Flashback: Collective Numerals and PT Nouns in Latin*

The last observation gives us the opportunity for a brief comparative discussion: in fact, the occurrence of collective numerals, both semantically contrastive and morphosyntactically selected (at least apparently, e.g., with PT nouns) is not limited to Slavic but occurs in other branches of Indo-European, including Latin (see the comparative study by Brugmann, 1907: 49), as well as in other language families: Ojeda (1997: 161–166) reviews relevant data from Finnish, Mongolian and Greenlandic.

For Latin, we have mentioned in (4) the occurrence of the plural form of the numeral *ūnus* 'one' with PT nouns. For numerals from '2' on, alongside cardinal numerals, Latin had inherited from PIE a series of collective numerals: *bīnī* '2', *trinī* '3', *quaternī* '4', *quīnī* '5', etc. Latin grammars report that these are selected with PT nouns, and this usage is widely documented in Latin texts.

- (53) *Inter bin-a castr-a Pompei atque Caesaris*
 between two-ACC.N.PL camp(N)-ACC.PL Pompey's and Caesar's
unum flumen tantum intererat Apsus (CAES., B.G. 3.19)
 one river only intervene:IPF-3SG Apsus
 'the two camps, Pompey's and Caesar's, were separated only by the river Apsus'

That this selection may have been obligatory seems to be suggested by passages by ancient grammarians such as the following.

- (54) non dicimus biga una, quadrigae duae, nuptiae tres, sed pro eo unae bigae, binae quadrigae, trinae nuptiae (VARR., *L.L.* 10.3.30)
 ‘we do not say **biga una*, **quadrigae duae*, **nuptiae tres* [with cardinal numerals] but instead say *unae bigae* “one two-horse team”, *binae quadrigae* “two pair of four-horse teams”, *trinae nuptiae* “three sets of nuptials” [with distributive numerals]’ (translation after Taylor, 1974: 96)

In mentioning the selection of *unae*, *binae*, *trinae* in (54), Varro voices grammatical prescription, recurring in the grammars from the antiquity. Slightly different statements are met with in Flavius Caper, 2nd century AD (Keil, 1856–1880: 7.108.7f.): “*binas tabulas dicimus, non duas*” ‘we say *binas tabulas* ‘two writing tablets’, not *duas*’; or Priscian, the most influential grammarian of Late antiquity (see Keil, 1856–1880: 3.414.25). But whether Varro’s and his fellow grammarians’ “*non dicimus*” can be taken as grammaticality judgements is dubious, in view of the fact that cardinal numerals are also attested with the same PT nouns (55), and even reported in the context of a metalinguistic observation by another grammarian, as is the case in Servius’ commentary on Vergil in (56).

- (55) a. *castra duo praesidiaria Barzalo et Claudias peterentur, sese ductante*
 ‘that they should go under his guidance to the two garrison camps of Barzalo and Claudiae’ (AMMIANUS MARCELLINUS, *Res gestae* 18.7.10, translation after Rolfe, 1982: 1.455)
- b. *exinde duabus admotis quadrigis in currus earum distentum inligat Mettium*
 ‘thereupon, two chariots were brought up, and he [Tullus Hostilius] orders that Mettius be stretched out between them’ (LIV. 1.28.10)
- (56) *frenaque bina poetice, nam ‘duo’ debuit dicere: ‘bina’ enim secundum Ciceronem non dicuntur nisi de his quae sunt numeri tantum pluralis. nam Cicero per epistolam culpavit filium, dicens male eum dixisse ‘direxi litteras duas’, cum ‘litterae’, quotiens epistolam significant, numeri tantum pluralis sint. contra ‘epistolas binas’ non dicimus, sed ‘duas’* (SERV. A. 8.168)

frenaque bina ‘two brakes’ is used in poetry, indeed he should have said *duo*: in fact *bina* according to Cicero can be used exclusively with PT nouns, since Cicero in a letter finds fault with his son, telling him that he made a mistake in saying *direxi litteras duas* ‘I sent two letters’, because *litterae*, whenever it means ‘letter’, is a PT noun. On the contrary, we do not say **epistolas binas*, but rather *epistolas duas* ‘two letters’ (compare *litteras duas/*binas* ‘two letters of the alphabet’)

Ammianus was a native speaker of Greek, born in Antioch in 330 AD, who learned Latin as L2 (Rolfe, 1982: 1.xx), but this was not the case for Livy nor Cicero’s son, Marcus Minor, whom his father rebuked according to Servius’ passage for saying, “incorrectly”, *litteras duas*.²² Based on this evidence, Löfstedt (1958: 101) argues that the use of collective numerals (which he labels ‘distributive’ following a tradition that goes back to the ancient grammarians: *disper-titiva* ‘distributives’ in Priscian, *De figuris numerorum*, ed. Keil, 1858: 3.413.24) with those nouns was determined by the semantics, and hence did not really differ from the occurrence of the same collective numerals with count nouns to count “Einheiten, deren jede in sich ein Mehrfaches ist” [‘units, each of which is per se a multiplicity’] (Löfstedt, 1958: 100). This latter use with count nouns is exemplified in the following examples (discussed in Ojeda, 1997: 146f.):²³

- (57) a. *molas asinarias unas et trusatilis unas Hispaniensis unas* (CATO, *agr.* 10.4)
 ‘one pair of donkey mills, one pair of hand mills, one pair of Spanish mills’
- b. *habetis interim bina animalia* (APUL., *De deo Socratis* 4)
 ‘you have in the meantime two kinds of creatures’
- c. *boves trinos* (CATO, *agr.* 10.1)
 ‘three yoke of oxen’

22 These examples have been discussed in many studies, from Brugmann (1907: 49 n. 1), who recognizes that the grammarians’ rule did not (any longer) mirror actual usage in Classical Latin, to Ojeda (1997: 154).

23 This emerges from Löfstedt’s (1958: 101) account of the occurrence of cardinal numbers in (55)–(56): “Die Verwendung von Kard. für Distr. in solchen Fällen erklärt sich wenigstens zum Teil dadurch, dass man das Gefühl verloren hatte, dass es sich um pluralische Einheiten handele; *litterae* war nicht mehr eine Gruppe von Buchstaben, sondern ein Brief, eine *epistula*.” [“The use of cardinal instead of distributive [i.e., collective] numerals is at least in part explained by the fact that one had lost the sense that these were *plural* units: *litterae* was no longer a group of letters, but a letter, an *epistula*.”]

In conclusion, a difference between (4) and (53) emerges: with PT such as *castra*, the plural form of *ūnus* was mandatory, while **unum castra* would have been ungrammatical, whereas the selection of collective *bīna, trina* (instead of *duo, tria*) with nouns of the same kind was optional.²⁴

This comparison corroborates the conclusion that the NIR replica numerals *dvoje, troje* selected categorically with the feminine PT nouns in (14), have unique properties. Their contact source is collective numerals whose semantics is still visible in NIR in their marginal use with count nouns exemplified in (48)–(52). However, categorical selection in, say, *dvoje/troje novine* ‘two/three newspapers’ is dictated by the morphosyntax, not by the semantics. In other words, these borrowed forms have become fully integrated in the NIR lower numeral lexemes filling a morphosyntactically defined paradigm cell, as shown in (45).

8.5 *Complicating Gender or Number? Comparative Evidence from Romance and Beyond*

When analysing rather intricate systems, ascribing a given contrast to one or the other morphosyntactic feature may prove a non-trivial issue. For instance, in his discussion of PT nouns Corbett (2019: 54f.) mentions Cicipu, a Benue-Congo language spoken in northwest Nigeria, in which there is just one PT, the noun *à-húlá* ‘name’, which “has a plural form, plural agreements, and this is so whether it denotes one name or more than one”; he adds in a footnote: “McGill (2009: 253) treats this noun as belonging to an inqorate gender, but I believe it should be seen as a number problem (it lacks a singular form) rather than a gender problem.” Similar problems present themselves also in Romance, and briefly addressing some of this evidence will help consolidating our analysis of NIR.

8.5.1 A Controversial Case: Asturian *o*-Agreement as a Value of Gender or Number

A case in point from Romance is that of (Central) Asturian, where all pronominal modifiers, exemplified in (58) with the definite article, mark the usual binary contrast (as in Spanish or Italian), while other agreement targets not

24 A comparable optionality is reported by Stefanović (2011) for contemporary BCS usage, as mentioned in Section 5 while commenting on (18)–(19). Other Slavic branches show a rather intricate situation. In Russian, a few PT nouns still select collective numerals categorically: e.g., *dvoe časóv* ‘two watches’ is the only grammatical way to quantify the PT noun *časý* ‘watches’ with a one-word numeral expression, while the cardinal numerals *dva* ‘two.M/N’/ *dve* ‘two.F’ are barred. Of course, paraphrase with a periphrastic classifier is always a viable alternative, which indeed seems the favourite one for several of the subjects tested by Nikunlassi (2000: 235–241).

preceding the noun within the NP signal a three-way distinction (data from the Central Asturian dialect of Lena; see Neira Martínez, 1955: 70–72; 1978: 260; the standardized variety of Asturian displays the same behaviour):

(58)	gender	countness	det	N	Adj	Central Asturian (Lena)
	F	count	<i>la</i>	<i>casa</i>	<i>fria</i>	‘DEF.F.SG cold house’
		mass		<i>tsiche</i>	<i>frio</i>	‘DEF.F.SG cold milk’
	M	mass	<i>el</i>	<i>café</i>	<i>frio</i>	‘DEF.M.SG cold coffee’
		count		<i>pie</i>	<i>friu</i>	‘DEF.M.SG cold foot’

This three-way distinction has been dubbed one of subnumber by Corbett (2000: 126), who proposes that the singular subdivides into mass and singular in a second number system:

(59)	Top system (number)	singular		plural
	Second system (subnumber)	mass	singular	plural
	(exemplified with masculine)	<i>fri-o</i>	<i>fri-u</i>	<i>fri-os</i>
				‘cold’

The alternative analysis proposed in Loporcaro (2018: 172–179), on the contrary, regards the binary contrast seen in (58) on definite articles and the three-way one seen on postnominal adjectives as manifestations of two concurrent gender systems, along the lines of the cross-linguistic study by Fedden and Corbett (2017).

(60) Alternative analysis (Loporcaro 2018: 172–192): two concurrent gender systems

SYSTEM 1: TWO VALUES			SYSTEM 2: THREE VALUES		
MASC.	<i>el pie/café</i>		MASC.	<i>pie</i>	<i>fri-u</i>
	DEF.M.SG		COUNT	foot(M,MC)	cold-MC
	foot(M,MC)/coffee(M,N)		NEUTER	<i>tsiche/café</i>	<i>fri-o</i>
				milk(F,N)/	cold-N
				coffee(M,N)	
FEM.	<i>l-a</i>	<i>casa/tsiche</i>	FEM.	<i>casa</i>	<i>fri-a</i>
	DEF-F.SG	house(F,FC)/	COUNT	house(F,FC)	cold-FC
		milk(F,N)			
DOMAIN: [__ N] _{NP}			DOMAIN: elsewhere		

In our NIR case, an analysis differing from the one put forward here in terms of (sub)gender and positing a number contrast instead, seems to be less likely, given the overall morphosyntactic structure of a Romance language like IR as well as the general properties of grammatical number. The values of the number feature are defined semantically in terms of the numerosity of real world entities.

These values of the number feature have meanings and forms associated with them. The main part of the meaning of the singular is that it refers to *one* real world entity, while the plural refers to *more than one* distinct real world entity. [emphasis added, *M.L. et al.*] (CORBETT, 2000: 4)

In the data in (37), the number value of all the contrasting items *doj/do/dvoje* and *trej/troje* is identical in terms of real world entities: the quantified NPS *dvoje novine/vile* ‘two newspapers/pitchforks’, *do mære/ženske* ‘two hands/women’, and *doj dints/omir* ‘two teeth/men’ all denote exactly two real world entities, and the same identity goes for *trej/troje*, so that there seems to be no cogent semantic/referential reason to postulate any contrast among them, as to this category. In Romance, where the number contrast is binary (singular vs. plural), quantified phrases containing ‘two’ and ‘three’ are all equally non-singular, i.e., plural. Alternatively, such a reason could be provided by the morpho-syntactic system, as is the case in languages such as Finnish.

8.5.2 A Different Case: Number Contrasts in Numerals in Finnish

Finnish shows “an unusual interaction between numerals and nouns”, thoroughly discussed in Hurford (2003: 584–589; quote from p. 584). In this language, all numerals have both singular and plural forms, the latter used to indicate sets of objects (contrast (61b) with the singular forms (61a)) and also selected obligatorily with PT nouns (61c) (Hurford, 2003: 587):²⁵

(61) a.	<i>yksi</i>	<i>kenkä</i>	/	<i>kaksi</i>		<i>sukaa</i>
	one:NOM.SG	shoe:NOM.SG		two:NOM.SG		sock:PRTV.SG
	‘one shoe’			‘two socks’		

25 In Finnish, plural numerals agree with head nouns in all cases. In the singular, this happens with *yksi* ‘one’, while other formally singular numerals govern a noun in the partitive singular, whenever the relevant NP receives nominative or accusative case, the only two cases occurring in (61)–(62) (Hurford, 2003: 585). In NPS which receive any of the remaining cases, case-agreement is observed.

- b. *yhdet* *kengät* / *kahdet* *sukat*
 one:NOM.PL shoe:NOM.PL two:NOM.PL sock:NOM.PL
 ‘one group (typically a pair) of shoes’ / ‘two groups (pairs) of socks’
- c. *kahdet* *sakset* / *kahdet* *kasvot*
 two:NOM.PL scissors:NOM.PL two:NOM.PL face:NOM.PL
 ‘two pairs of scissors’ / ‘two faces’

This is interesting in many respects, for our discussion. One reason is that, for NPs in which plural numerals modify count nouns, Hurford (2003: 588) describes diverging judgements among his informants, in a way somewhat reminiscent of the variation in interpretation discussed in (49)–(52) while commenting on what we have labelled the semantic use of borrowed *dvoji*, *dvoje* in NIR:

- (62) a. *oppilaat* *saivat* *kolme* *kirjaa*
 pupils got three:ACC.SG book:PRTV.SG
 ‘pupils received three books’
- b. *oppilaat* *saivat* *kolmet* *kirjat*
 pupils got three:ACC.PL book:ACC.PL
 ‘pupils received piles of three books’/‘three groups of pupils received (some) books’

Sentence (62a), containing a singular numeral, is systematically ambiguous for all informants – just as its English translation equivalent – between a reading where the quantified NP has wide scope (“there is a set of just three books which the pupils, as a group, receive”) and a distributive reading where *oppilaat* has scope over *kolme kirjaa* (“each individual pupil receives a set of three books”). Hurford’s (2003: 588) informants part ways when it comes to interpreting (62b), where, the plural numeral induces different interpretations: for one informant, “each pupil receives copies of the same three books as the other pupils”, while for another the reading is that “a teacher has three variously sized groups of pupils and gives each group of pupils one pile of books; we don’t know how many books are in each pile, but there are exactly three piles”. As Hurford (2003: 588) puts it, “What is common to the interpretations suggested by both informants is the idea of three sets (alias types, piles) of books.”

This very variation shows that number contrasts in numerals, though well-entrenched in the morphology and morphosyntax of Finnish, fall in a grey zone: while the unmarked option has an unambiguous meaning, the other one (here, the plural) is trickier. With count nouns, there is vacillation in interpretation and, in addition, Hurford (2003: 587) reports judgements by speakers who deem

plural numerals awkward in this or that context. With PT nouns, by contrast, the use of plural numerals (to the exclusion of singular ones) is described as categorical and unproblematic, and this is generally the case in Finnish grammars (cf. e.g., Whitney, 1956: 173). Thus, the Finnish evidence shows that a difference in number is an option, cross-linguistically, for numeral quantification with PT vs. plain count nouns. Finnish is well equipped for this, as its numerals are “declined in the same way as nouns” (Whitney, 1956: 171). By contrast, IE languages such as Latin and the Slavic languages take this option only for the numeral ‘one’, and even for this, many Romance languages – with the exceptions seen in (5), (8d) and (9) – have to resort to the classifier strategy instead. On the whole, thus, the Romance languages differ from Finnish in that they do not feature a declensional paradigm of numerals in which a regular number contrast can be hosted. Consequently, the distinction introduced by *dvoje* and *troje*, contrasting respectively with *doj/do* and *trej*, is doomed to remain an isolated irregularity, which is indeed what overdifferentiation means. When it comes to labelling the morphosyntactic feature involved, gender seems the natural choice in terms of system-adequacy, given the non-availability of number (contrary to Finnish) and given comparable cases of gender overdifferentiation on lower numerals in Romance (see Sections 8.1, 8.5.3).

The occurrence of minor number values, with restricted range in the lexicon of some languages (see Corbett, 2000: 89–110), might be described as a kind of pendant to gender overdifferentiation: for instance, in Arapesh (Papua New Guinea) “pronouns and nouns typically distinguish singular and plural [...]. But just the first person pronoun has singular versus dual versus plural” (Corbett, 2000: 91). Corbett’s cross-linguistic review of minor numbers does not include any examples from the Romance languages.²⁶

8.5.3 A Bipartite Gender Value (for One Class of Targets) in Northern Apulia

Thus far, we have argued that the contrast between *dvoje* and *do* must be accounted for in the morphosyntax rather than being a matter of mere semantics (Sections 8.3–8.4), and that an account in terms of (sub)gender seems preferable over one in terms of number (Sections 8.1–8.2; 8.5.1–8.5.2). As a final

26 Another way of treating systems where number does not behave uniformly across word classes is the distinction of a top and a second system (as shown in (59)), which can coexist with distinctions in range. Corbett (2000: 92f, 120f.) illustrates this point with Yimas (Papua New Guinea), in which both nouns and pronouns contrast singular, dual and plural, while only personal pronouns contrast paucal in addition. The additional contrast for this minor number value defines at the same time the top number system, covering personal pronouns, while the second system covers nouns.

piece of comparative evidence in support of our analysis, we will now show that there are indeed comparable cases of Romance varieties in which just one gender value is subdivided in two subgenders, contrasted on just one overdifferentiated agreement target. One such variety, the Northern Apulian dialect of Sannicandro Garganico (province of Foggia), is discussed in Loporcaro (2018: 289–291), based on data from Carosella (2005: 89) and Gioiosa (2000: 91–95). In Sannicandrese, only one class of targets, demonstratives, is sensitive to a [\pm human] contrast, and this sensitivity is restricted to the masculine (63a–b), one of the two gender values normally contrasted in the dialect, which shows elsewhere (on articles, adjectives, participles etc.) a plain binary contrast:²⁷

- (63) a. *kwidd-u* *krəstja:nə* Sannicandro Garganico
 DEM.DIST\M_HUM-M.SG man(M_HUM) (province of Foggia, Italy)
 ‘that man/person’
- b. *kwedd-u* *kavətsə:nə*
 DEM.DIST\NONM_HUM-M.SG trousers(M_NONHUM)
 ‘that pair of trousers’
- c. *kwedd-a* *kumma:ra/vakka/kami:ša*
 DEM.DIST\NONM_HUM-F.SG godmother/cow/shirt(F)
 ‘that godmother/cow/shirt’

More precisely, as specified in the glosses in (63), affixal inflection encodes the same binary masculine vs. feminine contrast found elsewhere, and it is only the combination of affixes with the allomorphs of the demonstrative stem that marks the subgender contrast: the allomorphs *kwidd-* (distal), *kwiss-* (intermediate), and *kwist-* (proximal, exemplified in (64)) occur with [masculine, singular, human] head nouns, while the complementary allomorphs *kwedd-*, *kwess-*, and *kwest-* occur elsewhere, including with [masculine, singular, non-human] head nouns – as shown in (64), Sannicandrese has a convergent system (Corbett, 1991: 155) neutralizing gender in the plural:

27 Note that *kavətsə:nə* in (63b) is a count noun and occurs there in the singular, just as the word *parrotsə* ‘black bread’ in (64): the corresponding plural(s) would have selected the plural form of the demonstrative, viz. *kwidd-i*.

(64)		SINGULAR	PLURAL	gloss
M	HUM	<i>u/kwist-u kumba:rə</i>	<i>i/kwist-i ġġú:vənə</i>	'DEF/DEM_PROX godfather/ youngster, -s'
	NONHUM	<i>u/kwest-u parrottsə</i>	<i>i/kwist-i kavaddə</i>	'DEF/DEM_PROX black bread/ horse, -s'
F		<i>la/kwest-a kami:ša</i>	<i>i/kwist-i kami:šə</i>	'DEF/DEM_PROX shirt, -s'

This parallel supports the analysis proposed for NIR lower numerals in (45), in that it shows that overdifferentiation within just one gender value on just one agreement target may arise anew, in a Romance variety.

9 Conclusion

The NIR case departs from the other cases of gender overdifferentiation in Romance discussed thus far, because neither in Romansh and medieval Northern Italo-Romance (40)–(43), nor in the Northern Apulian dialect mentioned in (63)–(64), was this overdifferentiation induced by contact. The two cases considered for comparison differ from each other, in turn, in that in Romansh both form and function (of e.g., Sursilvan *dus*, *duas*, and *dua*) are inherited (though the functional domain of *dua* has shrunk massively), whereas in Sannicandrese the forms are inherited but the functions have been reshuffled, since *kwist-u* vs. *kwest-u*, nowadays both masculine contrasting as [+human] vs. [–human], must be traced back to Late Latin masculine *ECCUM-ISTUM vs. neuter *ECCUM-ISTOC, i.e., to a gender contrast, not one of subgender.

In NIR, overdifferentiation in lower numerals arose via borrowing of *dvoje* and *troje* as a net increase in complexity (number of contrasts), thus adding to the not too many cases reported so far of contact-induced morphosyntactic complexification. On the whole, the NIR system has become more complex through contact in several ways, all involving borrowing from Croatian of agreement targets which had different functions in the source language. The symmetrically defective values of the two neuters (N₁ and N₂) both derive from one and the same non-defective gender value of Croatian, the neuter. The overdifferentiation on 'two' and 'three', by contrast, arose capitalizing on borrowed numeral forms which, in the source system, contrasted in lexical/semantic terms with non-collective numerals but, once borrowed, entered one and the

same lexeme paradigm with the Daco-Romance inherited numerals *doj/do* and *trej* respectively. This borrowing may have started as whole Croatian NPs headed by PT nouns and consequently containing collective numeral forms came to be used in NIR discourse, much like in the case of other numerically quantified borrowed NPs considered in Section 4. Also, this borrowing process cumulated onto another, also contact-induced, distinctive property of IR, viz. the availability of the F.PL form of the numeral *ur/ura* ‘one:M/F’ for quantification of PT nouns, seen in (5). This was probably a calque on Slavic, shared by NIR and SIR, which however did not in itself impact on gender since the F.PL form *ure*, selected with PT nouns, contrasted with *ur/ura* in number. By contrast, as *dvoje* and *troje* became novel forms in the paradigm of the numerals ‘two’ and ‘three’, adding to inherited *doj/do* and *trej*, a contrast in number was not an option, since all these forms are uniformly plural. This resulted in the subgender contrast we have described.²⁸

To sum up, the result of our analysis of NIR can be schematized as in (65), where the class one adjective *bur* ‘good’ illustrates the core grammatical system, originally consisting of the four inherited cells occupied by *bur*, *-a*, *-i*, *-e*. In addition, the paradigm of agreement targets such as *bur* has been enriched with the N₁ (*buro*), which found its way into the gender system (of both branches of IR), in spite of its scantiness in terms of controller lexemes, because of its syntactic function as the default agreement marker. At a later – and quite recent – stage, only in NIR the N₂ (*bura*) has arisen: this completes the set of agreement options available in today’s NIR for all class one adjectives, articles, personal pronouns and demonstratives. In addition, the paradigms of the two numerals ‘two’ and ‘three’ show the further complexification of the gender system in this Romance variety.

(65)

	M	F		N ₁	N ₂	
		collective				
SG	<i>bur</i>	<i>bura</i>		<i>buro</i>		‘good’
PL	<i>huri</i>	<i>bure</i>			<i>bura</i>	
	<i>doj</i>	<i>do</i>	<i>dvoje</i>		–	‘two’
	<i>trej</i>		<i>troje</i>		–	‘three’

28 Once the latter was established, also *ure* can be viewed as a form filling the now available collective F.PL subgender cell.

As we have argued, borrowing of *dvoje* and *troje* from Croatian, now selected categorically in NIR with a handful of feminine PT nouns, has enriched the paradigm of the two numeral lexemes at issue, but also affected the morpho-syntactic system, yielding (sub)gender overdifferentiation within the feminine. This was the rather unexpected conclusion our analysis brought us to, considering that the original purpose of our fieldwork in Istria was an inspection of the numeral system of this highly attrited, endangered language.

Abbreviations

BCS	=	Bosnian-Croatian-Serbian
Eng.	=	Engadinian
Lat.	=	Latin
(N/S)IR	=	(Northern/Southern) Istro-Romanian
PT	=	<i>plurale tantum/pluralia tantum</i>
Sl.	=	Slavic
Srs.	=	Sursilvan

Acknowledgements

The paper is part of the research project “Linguistic morphology in time and space (LiMiTS)” (Sinergia [SNF CRSII1_160739]): the Swiss National Science Foundation (SNF) as well as the UFSP “Sprache und Raum (SpuR)” and the Romance Seminar of the University of Zurich are gratefully acknowledged for funding the fieldwork, which has been carried out in June 2017 and June 2018. We are indebted to Goran Filipi for making fieldwork possible through introducing us to our first Istro-Romanian informants who kindly devoted their time to us answering our questions: our heartfelt thanks go to Lučjano and Julijana Turković, Adrijana Gabriš, Cvetko Doričić, Robert Doričić, Željko Doričić and Davorka Stambulić (from Žejane), Boris Bako and Josip Glavina (from Šušnjeveca) and Silvano and Kristina Karlović (from Jesenovik), who shared their native intuitions with us, opening their homes to our curiosity for the linguistic treasure they hold. Thanks are also due to Stefano Cristelli, Melita Lajqi, Stefano Negrinelli, Andrija Petrović, Olivier Winistörfer, and Mario Wild for help with the fieldwork. Finally, we thank Grev Corbett, Sebastian Fedden, Silvana Vranić, Nikola Vuletić and two anonymous reviewers for comments on a previous draft. Parts of this research have been presented at the workshops

“Romance languages and the others: the Balkan Sprachbund”, Zurich, May 2018 and “Gender in heritage languages and grammar change”, Oslo, October 2019; at *Going Romance*, Utrecht, December 2018; at the University of Florence, February 2019. We thank the audiences for useful comments. Usual disclaimers apply.

References

- Adamou, Evangelia. 2013. Replicating Spanish *estar* in Mexican Romani. *Linguistics* 51(6): 1075–1105.
- Aikhenvald, Alexandra Y. 2002. *Language Contact in Amazonia*. New York: Oxford University Press.
- Aikhenvald, Alexandra Y. 2003. Mechanisms of change in areal diffusion: New morphology and language contact. *Journal of Linguistics* 39(1): 1–29.
- Arkadiev, Peter and Francesco Gardani. 2020. Introduction: Complexities in morphology. In Peter Arkadiev and Francesco Gardani (eds.), *The Complexities of Morphology*, 1–19. Oxford: Oxford University Press.
- Ascoli, Graziadio Isaia. 1861. *Colonie straniere in Italia. Studi critici*. Vol. 1, 37–85. Milan: Politecnico [repr. Sala Bolognese: Forni, 1980].
- Bickerton, Derek. 1981. *Roots of Language*. Ann Arbor: Karoma Publishers.
- Breu, Walter. 2013. Zahlen im totalen Sprachkontakt: das komplexe System der Numeralia im Moliseslavischen. *Wiener Slawistischer Almanach* 72: 7–34.
- Brugmann, Karl. 1907. *Die distributiven und die kollektiven Numeralia der indogermanischen Sprachen. Mit einem Anhang von Eduard Sievers: altnordisch tvenn(i)r, þrenn(i)r, fernir*. Leipzig: Teubner.
- Buchi, Eva. 2006. Wieviel Wortbildung, wieviel Morphologie verträgt die etymologische Forschung? Bemerkungen zur Beschreibung rumänischer Slavismen. In Wolfgang Dahmen, Günter Holtus, Johannes Kramer, Michael Metzeltin, Wolfgang Schweickard, and Otto Winkelmann (eds.), *Lexikalischer Sprachkontakt in Südosteuropa*. Romanistisches Kolloquium XII, 73–90. Tübingen: Narr.
- Candinas, Theo. 1982. *Romontsch sursilvan. Grammatica elementara per emprender igl idiom sursilvan*. Cuera: Ligia Romontscha.
- Candinas, Theo. 2009. *Ein Elsässer im Ersten Weltkrieg/In schuldau d'Alsazia en l'Emprema uiara*. Frauenfeld: Reinhold Liebig Verlag.
- Carosella, Maria. 2005. *Sistemi vocalici tonici nell'area garganica settentrionale fra tensioni diatopiche e dinamiche variazionali*. Rome: Nuova cultura.
- Combi, Carlo. 1859. *Porta orientale*, Anno III. Fiume: Schubart – Rezza/Trieste: C. Coen.
- Corbett, Greville G. 1979. The agreement hierarchy. *Journal of Linguistics* 15(2): 203–224.
- Corbett, Greville G. 1983. *Hierarchies, Targets and Controllers. Agreement Patterns in Slavic*. London and Canberra: Croom Helm.

- Corbett, Greville G. 1991. *Gender*. Cambridge: Cambridge University Press.
- Corbett, Greville G. 2000. *Number*. Cambridge: Cambridge University Press.
- Corbett, Greville G. 2006. *Agreement*. Cambridge: Cambridge University Press.
- Corbett, Greville G. 2007. Deponency, syncretism and what lies between. In Matthew Baerman, Greville G., Dunstan Corbett Brown, and Andrew Hippisley (eds.), *Deponency and Morphological Mismatches*, 21–43. Oxford: British Academy and Oxford University Press.
- Corbett, Greville G. 2012. *Features*. Cambridge: Cambridge University Press.
- Corbett, Greville G. 2019. Pluralia tantum nouns and the theory of features: A typology of nouns with non-canonical number properties. *Morphology* 29(1): 51–108. DOI 10.1007/s11525-018-9336-0.
- Dahmen, Wolfgang and Johannes Kramer. 1976. Observații despre vocabularul istroromânei vorbite la Jeišn. *BANF* 1: 79–89.
- Dahmen, Wolfgang and Johannes Kramer. 1988. Le inchieste istro-rumene di Ugo Pellis. Parte prima: Questioni 1-1512. *Balkan-Archiv* (Neue Folge) 13: 209–281.
- de Groot, Casper. 2008. Morphological complexity as a parameter of linguistic typology: Hungarian as a contact language. In Matti Miestamo, Kaius Sinnemäki and Fred Karlsson (eds.), *Language Complexity: Typology, Contact, Change*, 191–215. Amsterdam and Philadelphia: John Benjamins.
- Despić, Miloje. 2017. Investigations in mixed agreement: Polite plurals, hybrid nouns and coordinate structures. *Morphology* 27(3): 253–310.
- Everett, Caleb. 2017. *Numbers and the Making of Us. Counting and the Course of Human Cultures*. Cambridge, Mass.: Harvard University Press.
- Fedden, Sebastian and Greville G. Corbett. 2017. Gender and classifiers in concurrent systems: Refining the typology of nominal classification. *Glossa: A Journal of General Linguistics* 2(1): 34. 1–47.
- Filipi, Goran. 2002. *Istrorumunjski lingvistički atlas. Atlasul Lingvistic Istroromân. Atlante Linguistico Istroromeno*. Pula: Znanstvena udruga Mediteran.
- Frățișă, Vasile and Gabriel Bărdășan (2010). *Dialectul istroromân. Straturi etimologice. Partea I*. Timișoara: Editura Universității de Vest.
- Ganzoni, Gian Paul. 1977. *Grammatica ladina. Grammatica sistematica dal rumantsch d'Engiadin' Ota per scholars e creschieus da lingua rumauntscha e tudas-cha*. Samedan: Lia Rumantscha.
- Ganzoni, Gian Paul. 1983. *Grammatica ladina. Grammatica sistematica dal rumantsch d'Engiadina Bassa per scholars e creschiüts da lingua rumantscha e francesa*. Samedan: Lia Rumantscha.
- Gardani, Francesco. 2008. *Borrowing of Inflectional Morphemes in Language Contact*. Frankfurt am Main: Peter Lang.
- Gardani, Francesco. 2012. Plural across inflection and derivation, fusion and agglutination. In Lars Johanson and Martine I. Robbeets (eds.), *Copies versus Cognates in Bound Morphology*, 71–97. Leiden and Boston: Brill.

- Gardani, Francesco. 2018. On morphological borrowing. *Language and Linguistics Compass* 12(10): 1–17.
- Gardani, Francesco. 2020. Morphology and contact-induced language change. In Anthony Grant (ed.), *The Oxford Handbook of Language Contact*, 96–122. Oxford: Oxford University Press.
- Gioiosa, Matteo. 2000. *Grammatica del dialetto sannicandrese*. San Nicandro Garganico: Gioiosa.
- Giudici, Alberto and Chiara Zanini. 2021. A plural indefinite quantifier on the Romance-Slavic border. *Word Structure* 14: 195–225.
- Givón, Talmy. 1979. Prolegomena to any sane creology. In Ian F. Hancock (ed.), *Readings in Creole Studies*, 3–35. Amsterdam and Philadelphia: John Benjamins.
- Gusmani, Roberto. 1979. Sull'induzione di morfemi. In Gerhard Ernst and Arnulf Stefenelli (eds.), *Sprache und Mensch in der Romania. Heinrich Kuen zum 80. Geburtstag*, 110–116. Wiesbaden: Steiner [repr. in *Saggi sull'interferenza linguistica*, 2nd edn. Firenze: Editrice Le Lettere, 1986: 155–164].
- Haspelmath, Martin and Uri Tadmor (eds.). 2009. *Loanwords in the World's Languages: A Comparative Handbook*. Berlin: De Gruyter Mouton.
- Hockett, Charles F. 1958. *A Course in Modern Linguistics*. New York: Macmillan.
- Hurford, Jim. 2003. The interaction between numerals and nouns. In Frans Plank (ed.), *Noun Phrase Structure in the Languages of Europe*, 561–620. Berlin and New York: Mouton De Gruyter.
- Hurren, Anthony H. 1969. Verbal aspect and archi-aspect in Istro-Rumanian. *La Linguistique* 5(2): 59–90.
- Hurren, Anthony H. 1999. *Istro-Rumanian: A Functionalist Phonology and Grammar*. PhD dissertation, University of Oxford.
- Kalsbeek, Janneke 1998. *The Čakavian Dialect of Orbanici near Žminj in Istria*. Amsterdam and Atlanta: Rodopi.
- Kämpf, Sylvina. 2015. *Der Kollektiv in der Romania – eine Untersuchung des Bündnerromanischen*. BA thesis, University of Zurich.
- Karlsson, Fred, Matti Miestamo and Kaius Sinnemäki. 2008. Introduction: The problem of language complexity. In Matti Miestamo, Kaius Sinnemäki and Fred Karlsson (eds.), *Language Complexity: Typology, Contact, Change*, vii–xiv. Amsterdam and Philadelphia: John Benjamins.
- Keil, Heinrich. 1856–1880. *Grammatici Latini*. 7 vols. Leipzig: Teubner [repr. Hildesheim: Olms, 2007].
- Kim, Hyoung Sup. 2009. *The structure and use of collective numeral phrases in Slavic: Russian, Bosnian/Croatian/Serbian, and Polish*. PhD dissertation, University of Texas at Austin.
- Klockmann, Heidi. 2017. *The Design of Semi-lexicality: Evidence from Case and Agreement in the Nominal Domain*. Utrecht: LOT.

- Kossmann, Maarten. 2010. Parallel System Borrowing: Parallel morphological systems due to the borrowing of paradigms. *Diachronica* 27(3): 459–487.
- Kovačec, August. 1963. Notes de lexicologie istroroumaine. Sur la disparition des mots anciens et leur remplacement par des mots croates. *Studia Romanica et Anglica Zagrabienisia* 15–16: 3–39.
- Kovačec, August. 1966. Quelques influences croates dans la morphosyntaxe istroroumaine. *Studia Romanica et Anglica Zagrabienisia* 21–22: 57–75.
- Kovačec, August. 1968. Observations sur les influences croates dans la grammaire istroroumaine. *La Linguistique* 1: 79–115.
- Kovačec, August. 1971. *Descrierea istroromânei actuale*. Bucharest: Editura Academiei Republicii Socialiste România.
- Kovačec, August. 1992. Éléments italiens du lexique istroroumain. *Linguistica* 32(2): 159–175.
- Kovačec, August. 1998. *Istrorumunjsko-Hrvatski Rječnik (s gramatikom i tekstovima)*. Pula: Znanstvena udruga Mediteran.
- Kusters, Wouter. 2003. *Linguistic Complexity: The Influence of Social Change on Verbal Inflection*. Utrecht: LOT.
- Leko, Nedžad. 2009. *The Syntax of Numerals in Bosnian*. Munich: Lincom Europa.
- Livescu, Maria. 2008. Histoire interne du roumain: morphosyntaxe et syntaxe/Interne Sprachgeschichte des Rumänischen: Morphosyntax und Syntax. In Gerhard Ernst, Martin-Dietrich Gleßgen, Christian Schmitt and Wolfgang Schweickard (eds.), *Romanische Sprachgeschichte/Histoire linguistique de la Romania. 3. Teilband*, 2646–2692. Berlin and New York: De Gruyter.
- Löfstedt, Bengt. 1958. Zum Gebrauch der lateinischen distributiven Zahlwörter. *Eranos* 56: 71–117, 188–223.
- Loporcaro, Michele. 2018. *Gender from Latin to Romance: History, Geography, Typology*. Oxford: Oxford University Press.
- Loporcaro, Michele and Lorenzo Tomasin. 2016. Marcamento di genere iperdifferenziato come vestigio del neutro nell'italo-romanzo settentrionale antico. *Lingua e Stile* 51: 37–64.
- Lučić, Radovan. 2015. Observations on collective numerals in Standard Croatian. *Journal of Slavic Linguistics* 23(1): 3–31.
- Małeck, Mięczyśław. 1930. *Przegląd słowiańskich gwar Istrji* [Overview of the Slavic Dialects of Istria]. Kraków: Polska akademja umiejętności.
- Martin, Samuel Elmo. 2004. *A Reference Grammar of Japanese*. Honolulu: University of Hawai'i Press [first published by Yale UP, New Haven and London 1975; revised edn. by Charles E. Tuttle Company, Rutland, Vermont/Tokyo, Japan 1988].
- Matras, Yaron. 2007. The borrowability of grammatical categories. In Yaron Matras and Jeanette Sakel (eds.), *Grammatical Borrowing in Cross-linguistic Perspective*, 31–74. Berlin and New York: De Gruyter.

- Matras, Yaron. 2009. *Language Contact*. Cambridge: Cambridge University Press.
- McGill, Stuart J. 2009. *Gender and Person Agreement in Cicipu Discourse*. PhD dissertation, SOAS, University of London.
- McWhorter, John H. 2001. The world's simplest grammars are creole grammars. *Linguistic Typology* 5(2–3): 125–166.
- McWhorter, John H. 2007. *Language interrupted. Signs of non-native acquisition in standard language grammars*. Oxford and New York: Oxford University Press.
- Meakins, Felicity and Sasha Wilmoth. 2020. Overabundance resulting from language contact: Complex cell-mates in Gurindji Kriol. In Peter Arkadiev and Francesco Gardani (eds.), *The Complexities of Morphology*, 81–104. Oxford: Oxford University Press.
- Melissaropoulou, Dimitra. 2017. On the role of language contact in the reorganization of grammar: A case study on two Modern Greek contact-induced varieties. *Poznan Studies in Contemporary Linguistics* 53(3): 449–485.
- Mihăilă, Gheorghe. 1960. *Împrumuturi vechi sud-slave în limba română. Studiu lexicosemantic*. Bucharest: Editura Academiei Republicii populare romîne.
- Mühlhäusler, Peter. 1980. Structural expansion and the process of creolization. In Albert Valdman and Arnold R. Highfield (eds.), *Theoretical Orientations in Creole Studies*, 19–55. New York: Academic Press.
- Neira Martínez, Jesús. 1955. *El habla de Lena*. Oviedo: IDEA [repr. 2005].
- Neira Martínez, Jesús. 1978. La oposición 'continuo'/'discontinuo' en las hablas asturianas. In *Estudios ofrecidos a Emilio Alarcos Llorach*, III, 255–279. Oviedo: Universidad de Oviedo [repr. in Id., *Bables y castellano en Asturias*. Madrid: Silverio Cañada, 1982: 163–186].
- Nikunlassi, Ahti O. 2000. The use of collective numerals in contemporary Russian: An empirical approach. *Wiener Slawistischer Almanach* 45: 209–246.
- Ojeda, Almerindo E. 1997. A semantics for the counting numerals of Latin. *Journal of Semantics* 14: 143–171.
- Payne, John and Rodney Huddleston. 2002. Nouns and noun phrases. In Rodney Huddleston and Geoffrey K. Pullum (eds.), *The Cambridge Grammar of the English Language*, 323–523. Cambridge: Cambridge University Press.
- Petrovici, Emil. 1962. Phonetic evolution, substitution of sounds or morphological accommodation? (In connection with the treatment of final -o in the Slavonic elements of Rumanian). *Revue roumaine de linguistique* 7: 17–20.
- Petrovici, Émile. 1967. Le neutre en istro-roumain. *To honor Roman Jakobson. Essays on the occasion of his seventieth birthday, 11 October 1966*, 1523–1526. The Hague and Paris: Mouton.
- Petrovici, Emil and Petru Neiescu. 1964. Persistența insulelor lingvistice. Constatări făcute cu prilejul unor noi anchete dialectale la istroromîni, meglenoromîni și aromîni. *Cercetări de Lingvistică* 9(2): 187–214.

- Petrovici, Emil and Petru Neiescu. 1965. Persistence des îlots linguistiques. Constatations faites à l'occasion de nouvelles enquêtes dialectales chez les istro-roumains, mégléno-roumains et aroumains. *Revue roumaine de linguistique* 10(4): 351–374.
- Pranjković, Ivo. 2000. *Hrvatski jezik. Udžbenik za 3. razred gimnazije*. Zagreb: Školska knjiga.
- Pușcariu, Sextil (în colaborare cu d-nii M. Bartoli, A. Belulovici și A. Byhan). 1906. *Studii istroromâne*. Vol. 1. *Texte*. Bucharest: Cultura națională.
- Pușcariu, Sextil (în colaborare cu M. Bartoli, A. Belulovici și A. Byhan), 1926. *Studii istroromâne*. Vol. 2. *Introducere, gramatică, caracterizarea dialectului istroromân*. Bucharest: Academia Română.
- Ribarić, Josip. 1940. *Razmještaj južnoslovenskih dijalekata na poluotoku Istri* [Distribution of South Slavic Dialects on the Istrian Peninsula]. Belgrade: Srpska Kraljevska Akademija.
- Rölfe, John C. (ed.). 1982. *Ammianus Marcellinus* [The Loeb Classical Collection; 1st edn. 1935]. Cambridge, Mass.: Harvard University Press and London: William Heinemann Ltd.
- Sala, Marius. 2013. Contact and borrowing. In Martin Maiden, John C. Smith and Adam Ledgeway (eds.), *The Cambridge History of the Romance Languages*. Vol. 2. *Contexts*, 187–236. Cambridge: Cambridge University Press.
- Schulte, Kim. 2009. Loanwords in Romanian. In Martin Haspelmath and Uri Tadmor (eds.), *Loanwords in the World's Languages: A Comparative Handbook*, 230–259. Berlin: De Gruyter Mouton.
- Seifart, Frank. 2015. Direct and indirect affix borrowing. *Language* 91(3): 511–532.
- Šipka, Danko. 2007. *A Comparative Reference Grammar of Bosnian, Croatian, Serbian*. Edited by R. David Zorc. Hyattsville, Md.: Dunwoody Press.
- Spescha, Arnold. 1989. *Grammatica sursilvana*. Cuera: Casa editura per mieds d'instrucziun.
- Stankiewicz, Edward. 1983. The collective and counted plurals of the Slavic nouns. In Michael S. Flier (ed.), *American contributions to the Ninth International Congress of Slavists*, Kiev, September 1983, Vol. 1. *Linguistics*, 277–292. Columbus, Ohio: Slavica Publishers [repr. in Edward Stankiewicz, *The Slavic Languages: Unity in diversity*, 173–170. Berlin: Mouton De Gruyter 1986].
- Stefanović, Aleksandar. 2011. Les numéraux en serbo-croate (bosniaque, croate, monténégrin, serbe). Normes des standards et problèmes syntaxiques. *Revue des études slaves* 82(4): 709–714.
- Stefanović, Aleksandar. 2014. Kvantifikacija imenica pluralia tantum. *Sarajevski filološki susreti II*, 45–67. Sarajevo: Bosansko filološko društvo.
- Stevanović, Mihailo. 1989. *Savremeni srpskohrvatski jezik. Gramatički sistemi i knjižnevojezička norma. I. Uvod. Fonetika. Morfologija* [Contemporary Serbo-Croatian

- Language. Grammatical Systems and Literary Norm. Vol. 1. Introduction. Phonetics. Morphology]. Belgrade: Naučna knjiga.
- Tagliani, Roberto. 2011. *Il Tristano corsiniano. Edizione critica*. Rome: Scienze e Lettere.
- Tagliavini, Carlo. 1972. *Le origini delle lingue neolatine*. 6th edn. Bologna: Pàtron.
- Taylor, Daniel J. 1974. *Declinatio. A Study of the Linguistic Theory of Marcus Terentius Varro*. Amsterdam: John Benjamins.
- ThLL = *Thesaurus linguae Latinae. Editus auctoritate et consilio academiarum quinque Germanicarum Berolinensis Gottingensis Lipsiensis Monacensis Vindobonensis*. Leipzig: Teubner 1900ff.
- Thomason, Sarah G. and Terrence Kaufman. 1988. *Language Contact, Creolization, and Genetic Linguistics*. Berkeley: University of California Press.
- Trask, Robert L. 1997. *A Student's Dictionary of Language and Linguistics*. London: Arnold.
- Trudgill, Peter. 2009. Sociolinguistic typology and complexification. In Geoffrey Sampson, David Gil and Peter Trudgill (eds.), *Language Complexity as an Evolving Variable*, 98–109. Oxford: Oxford University Press.
- Trudgill, Peter. 2011. *Sociolinguistic Typology: Social Determinants of Linguistic Complexity*. Oxford: Oxford University Press.
- Vallortigara, Giorgio and Nicola Panciera. 2014. *Cervelli che contano*. Milan: Adelphi.
- Vanhove, Martine. 2001. Contacts de langues et complexification des systèmes: le cas du maltais. *Faits de Langues* 18: 65–74.
- Vrzić, Zvezdana and Robert Doričić. 2014. Language contact and stability of basic vocabulary: Croatian words for body parts in Vlashki/Zheyanski (Istro-Romanian). *Fluminensia* 26(2): 105–122.
- Vuletić, Nikola. 2014. Les minorités linguistiques invisibles et/ou cachées de la Croatie: les communautés arbënishtë, istro-roumaine et istriote. In Ksenija Djordjević Léonard (ed.), *Les minorités invisibles: diversité et complexité (ethno)sociolinguistiques*, 182–192. Paris: Michel Houdiard Éditeur.
- Whitney, Arthur H. 1956. *Teach yourself Finnish*. London: The English Universities Press Ltd.