

Notes and updates on language contact between Sardinian and Italian/French: A corpus-driven analysis

Simone Pisano,^a Valentina Piunno,^b Vittorio Ganfi^c

^a Università per Stranieri di Siena, Siena, Italy <simone.pisano@unistrasi.it>

^b Università degli Studi di Bergamo, Bergamo, Italy <valentina.piunno@unibg.it>

^c Università degli Studi del Molise, Campobasso, Italy <vittorio.ganfi@unimol.it>

This contribution provides a description of language contact phenomena of heritage Sardinian in migration contexts. The analysis takes into account original data collected in the period 2008-2010 among Sardinian communities of immigrants in continental Italy and France. The paper particularly focuses on the outcome of contact phenomena in lexicon, syntax and pragmatics, in order to demonstrate how these levels of analysis operate in language contact and how strictly they interact. Among lexical features, this analysis takes into account the presence of: (i) integrated and adapted borrowings, (ii) the nature of borrowed items in terms of vocabulary stratification (e.g. specialised/technical terms vs more commonly used terms), (iii) collocations and multiword expressions. As far as syntactic features are concerned, the areas of linking words and pronominal structures introducing subordinate sentences are investigated. The field of information structure is also investigated through the analysis of the use of cleft sentences by Sardinian speakers living in France. As far as pragmatics is concerned, the role of discourse markers in language contact is discussed and their different functional classes are identified and examined.

KEYWORDS: Sardinian, contact phenomena, corpus-driven analysis.

1. *Contact phenomena. The case of Sardinian*

Sardinia's linguistic pluralism is a constant in the island's history: as many scholars claimed (Rindler Schjerve 1987, 1998, 2000, 2017; Blasco Ferrer 1984; Loi Corvetto 1992, 1993; Dettori 2002; Bolognesi & Heeriga 2005; Gargiulo 2009, 2011; Putzu 2011; Pisano 2015, 2017; Calaresu & Pisano 2017; Linzmeier 2019), the side-by-side presence of at least two languages (where one is usually the high pole of diglossia) has always characterised Sardinia's linguistic history and its society. However, the role of Sardinian was still dominant at least until the end of the 19th century, when Italian spread even to the lowest registers of communication. In this regard, Viridis noted that before the second part of 19th century "Sardinian was the language universally known and spoken by all social classes, and, although in a situation of practical diglos-

sia”, it kept “the status and dignity of a language, not yet reduced to a dialect” (2017: 169, translation ours).

Language contact phenomena have been thoroughly investigated since the crucial work by Weinreich (1953), and a number of analyses of several kinds of contact scenarios have been produced since the end of the 20th century (cf. Hickey 2010, and references therein).¹ Contemporary studies mainly agree on “the assumption that language contact is about the way in which linguistic systems influence one another. Contact-induced language change is consequently seen as change that is ‘external’” (Matras 2010: 66). The effects of language contact have been investigated by scholars interested in language typology, language history, dialectology and sociolinguistics² (Hickey 2010). Many studies have recently demonstrated that “essentially, any part of language structure can be transferred from one language to another” (Heine & Kuteva 2005: 1, and references therein) and have highlighted the influence of language contact on lexicon, morphology and phonology, as well as on grammatical features (cf., among others, Heine & Kuteva 2005, Aikhenvald 2006). Regarding Sardinian varieties, most studies on language contact have mainly focused on lexicon (see Wagner 1997; Paulis 1983, 2017; Dettori 2002, 2017; Barbatto 2017;³ Toso 2017; Viridis 2017). More recently, however, a further interest in syntax and morphology has characterised the research on language contact (between Sardinian and *superstratum* languages) and multilingual communication strategies (see Pinto 2012, 2015; Calaresu & Pisano 2017; Pisano 2017; Gaidolfi 2017a-b). In this paper we resort to the theoretical model listed above as a general framework to analyse language contact. Due to the nature of the phenomena investigated in this work (§3), the analysis of language contact will be supported by the theoretical models of cognitive linguistics, such as the Construction Grammar framework as well as the one of grammaticalisation studies.

Despite the growing interest in the Italian heritage languages (cf. Bettoni & Rubino 1996, Haller 1997, 2006, Turchetta 2005, Prifti 2013, among others), a complete account of varieties of Sardinia in migration contexts is still lacking. In particular, a corpus linguistic approach could be extremely useful to highlight the most recurrent features of language contact phenomena in the heritage scenarios. To deal with Sardinian as a heritage language, we follow Montrul (2016: 15) in considering heritage languages not only as immigrant languages (e.g. Spanish, Hindi, Russian, Korean, Mandarin, Arabic, and Tagalog in the United States, or Turkish and Polish in the Netherlands and Germany etc.) and aboriginal languages (e.g. Navajo in the United States, Inuttitut in Canada or Dyrirbal in Australia, etc.), but also national minority languages

(which may have official status, as is partially the case for Sardinian).⁴ It worth noting that Montrul (2016) includes in this group both languages belonging to a different language family from the official language of a country (e.g. Basque in Spain and France, Welsh in Wales, Greek and Aromanian in Albania) and languages of the same language family (e.g. Catalan in Spain, Frisian in the Netherlands and Germany).⁵

2. Contact phenomena in Sardinian: the method

Even though the regional Italian of Sardinia has gained considerable attention in the literature (Loi Corvetto 1983, 1992; Lavinio 1975, 2019; Gargiulo 2009, 2011, 2014; Piredda 2013), the way Sardinian has changed compared to Italian (e.g. lexical traits, morpho-syntactic structures, pragmatics, intonation) has not received enough attention yet. With only a few exceptions (cf. Gaidolfi 2017a-b), there is still much work to do in the field of contact phenomena between Sardinian and Italian. Furthermore, there is still no adequate documentation of Sardinian speech in general. Contact phenomena are particularly interesting in the minority language scenario, since they can reveal (i) how much a language – especially a minority language – is influenced by the hegemonic languages, and (ii) which traits of the hegemonic language are more frequently adopted by the minority language, thus which language features are most likely to change on the basis of language contact.

This analysis experiments qualitative and quantitative methods by using data extracted from a corpus of spoken texts. The investigation is based on a corpus-driven approach (Tognini-Bonelli 2001), through which findings are driven inductively by corpus data. As McEnergy & Gabrielatos (2006: 36) note,

[c]orpus-driven research aims at discovering facts about language free from the influence of existing theoretical frameworks, which are considered to be based on intuitions, and, therefore, are not comprehensive or reliable.

In the corpus-driven approach “theoretical statements are fully consistent with, and reflect directly, the evidence provided by the corpus” (Tognini Bonelli 2001: 84). We believe that this method may lead to a new perspective on contact phenomena in Sardinian: in this case, the corpus is not intended as a simple repository of examples, but as “the sole source of our hypotheses about language” (Tognini-Bonelli 2001: 84).

2.1. *The corpus*

The data analysed in this investigation have been previously obtained throughout the course of the project *Mànnigos de memòria in limbis dae su disterru* ('Tastes of memory in the languages of the diaspora'), carried out in 2008-2010.⁶ This project aimed at documenting the memory of Sardinians living in the Italian peninsula and in France belonging to networks of Sardinian associations. To this purpose, 210 video/audio interviews were conducted among Sardinian immigrants' communities in the Italian continent and in France.⁷ The project involved only native speakers of all the varieties of Sardinia, including Gallurese, Sassarese, Tabarchino and Catalan of Alghero.⁸ The interviews were conducted in the form of a 'semi-directive' conversation (i.e. a structured questionnaire which allows subjective digressions to the interviewed, cf. Grassi *et al.* 1997).

The corpus includes two subcorpora: (i) *FRA_Sard corpus*, which is related to the world of Sardinian emigration in France, and (ii) *ITA_Sard corpus*, concerning the Sardinian diaspora in the Italian peninsula. The contents of the two subcorpora are not comparable with each other and are kept separated, since they differ in terms of types/number of informants and number of tokens. Nevertheless, they are included in the dataset because they are representative of several different language contact phenomena with Italian and French, which are considered the roof languages of each sample of data.⁹

The 210 informants that have been interviewed have been divided into three macro-categories: (i) men over 35 years old, (ii) women over 35 years old, (iii) people of both sexes aged 18-35 years (first and second/third generation).¹⁰ All interviews deal with migration stories, even though other topics may arise according to the category of the interviewed (cf. Pisano *et al.* 2022 for the detailed description of the questionnaire).

The first and second categories include members, and/or executives and founders of Sardinian associations. They are particularly interesting from a cultural perspective: while the former category provides evidence of the changes that affected traditional societies since the 1960s, the latter group helps document reunification stories. The topics differentiation depending on the speaker's gender could seem rather odd; nevertheless, data were not collected following a linguistic perspective exclusively, since they have also been used for anthropological purposes. As is well known, female and male activities were traditionally distinguished, and this distinction is usually kept in emigration contexts, especially in the case of first generations. Indeed, this distinction is not kept for the youths group. The third group is very heterogeneous: it contains young people born and grown in Sardinia, as well as second/third gen-

eration Sardinians. In this case, the interviews concerned specific topics, depending on the subcategory to which the informant belongs.

In a preliminary stage, only 15% of the interviews has been transcribed. Since the majority of informants are native speakers of a Sardinian variety¹¹ (the informants of other minority languages of Sardinia are less represented in the dataset), the transcribed material only includes Middle-Southern and Middle-Northern Sardinian varieties. This means that all the examples listed in this paper only refer to Sardinian dialects.¹² As far as the transcription is concerned, the originality of the individual interviews from a lexical, syntactic and grammatical perspective has been kept. Thus, phenomena of interference, doubts or uncertainties of the informants have been included, as well as the linguistic peculiarities of the informant's variety. Data have been first computationally analysed through the Sketch Engine platform, an online tool for the automatic mark-up of texts and for the creation of electronic corpora. In particular, this tool allowed us (i) to identify/quantify the tokens of our corpus, and (ii) to perform a preliminary morphosyntactic annotation of each graphic form. To this end, the standard automatic annotation for Italian texts was applied.¹³ The set of transcribed texts consists of 100,806 tokens, and 953 sentences.

The corpus proves to be particularly useful for the analysis of several types of phenomena, including in particular consequences of language contact, which have been manually tagged in the corpus, according to their main features (e.g. specialised terms, discourse markers, etc.). Previously tagged contact phenomena have been then extracted, collected into different sets, and classified according to their lexical or functional features (e.g. lexical loans or calques, syntactic loans, borrowings playing a pragmatic function in both replica and model languages).

3. Data analysis

This study focuses on the examination of the outcome of contact phenomena in lexicon (§3.1), syntax (§3.2) and pragmatics (§3.3). In particular, the role of these levels of analysis and of their strict interaction in language contact is discussed.

3.1. Contact phenomena from the lexical viewpoint

The field of language contact has been widely analysed from the lexical point of view. Indeed, lexicon is one of the main fields where the outcome of code overlapping is evident. The integration of lexical material in one language from a model language typically “begins with

lexical units and only much later begins to affect grammatical units” (Siemund 2008: 5). Several lexical interference phenomena are attested in the corpus. Borrowing processes involve nouns, verbs, adjectives, adverbs, as well as functional units like prepositions and conjunctions. Borrowings are often integrally acquired, but a number of occurrences of adapted loans are also recorded in the corpus (§3.1.1).

This paragraph mainly focuses on linguistic data and their classification from a qualitative/quantitative perspective, while sociolinguistic features (such as origin, age, generation) of the informants are not discussed. Here we provide sociolinguistic information only when they are a distinctive feature for data classification.

For the purpose of this analysis, 747 borrowings have been identified in the corpus, in particular 623 in *ITA_Sard* corpus and 124 in *FRA_Sard* corpus. Table 1 summarises the number of borrowings for each subcorpus (in terms of type frequency), and distinguishes them according to their part of speech / function:

| | <i>ITA_SARD CORPUS</i> | <i>FRA_SARD CORPUS</i> | <i>TOTAL</i> |
|---|------------------------|------------------------|--------------|
| Nouns | 246 | 57 | 303 |
| Verbs | 101 | 2 | 103 |
| Adjectives | 82 | 17 | 99 |
| Adverbs | 132 | 37 | 169 |
| Prepositions | 12 | 3 | 15 |
| Conjunctions | 9 | 2 | 11 |
| Pronouns | 7 | 1 | 8 |
| Interjections, formulas, discourse markers | 34 | 5 | 39 |
| TOTAL | 623 | 124 | 747 |

Table 1. Total amount of borrowings identified in the corpus.

The class of borrowed nouns is the most represented in the corpus (40.56% of borrowings in terms of type frequency). Borrowed nominal lexemes include several terms of the ‘basic vocabulary’ (De Mauro 1999a-b) of Italian¹⁴, as well as terms of the ‘common vocabulary’ (1d), which includes the great majority of nominal loans in the corpus. Nouns belonging to specialised areas (1e) only rarely occur (at least as single words, cf. §3.1.2).

- (1) a. *maggioranza* ‘majority’, *nonna* ‘grandmother’, *babbo* ‘dad’, *papà* ‘dad’, *cucina* ‘kitchen’.
b. *prospettiva* ‘perspective’, *bronzo* ‘bronze’, *propaganda* ‘propaganda’, *cugina* ‘cousin’.
c. *manodopera* ‘manpower’, *ottone* ‘brass’, *balletto* ‘ballet’, *madrina* ‘godmother’.
d. *fisionomia* ‘physiognomy’, *alfabetizzazione* ‘literacy’, *andazzo* ‘style’, *frutteto* ‘orchard’.
e. *nullaosta* ‘permission’, *biomassa* ‘biomass’.

From the semantic viewpoint, it is worth noting the group of kinship terms, which are also characterised by a relatively high frequency of use. Some well-established loans – which may have entered in Sardinian from other languages through the influence of Italian – are also recorded (e.g. *babysitter*, *computer*, *fiction*, *lobby*, *residence*, *valzer*).

Verbal borrowings are less frequent with respect to nominal entities (13.79% of borrowings). They are lexically full verbs and mainly belong to the stratification of fundamental vocabulary (2):

- (2) *proporre* ‘propose’, *coinvolgere* ‘involve’, *salvaguardare* ‘safeguard’, *esprimere* ‘express’,
cantare ‘sing’, *ripetere* ‘repeat’, *riguardare* ‘concern’, *insegnare* ‘teach’, *trasferire* ‘move’.

Interestingly, verbal loans more often appear in the non-finite form. Indeed, more than half of verbal borrowings are used in the infinitive form; this form is structurally determined (they often appear after a modal verb, e.g. *si podet proporre* ‘it can be proposed’, *mi so dèvidu raccomandare* ‘I had to make a recommendation’, *pro sa cale nos devimus battere* ‘for which we have to fight for’, or in a verbal periphrasis *apo cumenzau a m’inserire* ‘I started including myself’), without any implications due to the dominant language (since the structures are normally present in Sardinian).

Adjectives represent a small class of borrowings (13.25%). The corpus highlights simplex (underived) qualitative adjectives (3a), as well as deverbal (3b) and relational (3c) ones (cf. Bosque & Picallo 1996: 351):

- (3) a. *enorme* ‘huge’, *qualsiasi* ‘any’, *migliore* ‘best’, *straordinaria* ‘extraordinary’, *piccolo* ‘small’.
b. *accogliente* ‘welcoming’, *gestibile* ‘manageable’, *ambita* ‘desirable’, *attiva* ‘active’.
c. *intercontinentale* ‘intercontinental’, *meritevole* ‘deserving’, *prestigiosa* ‘prestigious’.

Beyond being the second most represented class of borrowings in the corpus with 22.62% of exemplars, the class of borrowed adverbs is very interesting from the lexical and morphological points of view. Adverbial loans are mainly derived by means of the suffix *-mente*.

- (4) *economicamente* ‘economically’, *seriamente* ‘seriously’, *ininterrottamente* ‘uninterruptedly’,
veramente ‘really’, *interiormente* ‘inwardly’, *positivamente* ‘positively’, *naturalmente* ‘naturally’.

As already noted in Pisano *et al.* (2022: 158, 160), the use of *-mente* adverbs is strictly related to the need to fill a lexical gap in Sardinian,

which “does not have the same degree of lexical elaboration as the roof language” (Pisano *et al.* 2022: 160, translation ours). Adverbial loans involve only to a lesser extent simplex adverbs:

- (5) *soprattutto* ‘above all’, *purtroppo* ‘unfortunately’, *appunto* ‘precisely’, *ormai* ‘by now’, *intanto* ‘meanwhile’, *abbastanza* ‘quite’, *allora* ‘then’, *piuttosto* ‘rather’, *altrimenti* ‘otherwise’.

As better shown in §3.3., their use is mainly connected to pragmatic needs (e.g. they are often employed to highlight the speaker’s opinion).

Finally, it is possible to mention the use of borrowed prepositions (6), conjunctions (7) and pronouns (8), which constitute 2.01%, 1.47% and 1.07% of borrowings respectively:

- (6) *sotto* ‘under’, *tramite* ‘through’, *attraverso* ‘across’, *vicino* ‘near’, *dopo* ‘after’.
(7) *dopo* ‘after’, *nonostante* ‘after’, *cosicché* ‘so that’, *anche se* ‘even if’, *ciò* ‘that is’, *appena* ‘just’, *malgrado* ‘despite’, French *mais* ‘but’.
(8) *chiunque* ‘anyone’, *la quale* ‘which’.

With respect to other lexemes, borrowed function words are less common, even if their use is very frequent in the corpus. Interestingly, this could have a pragmatic reason: it may suggest that the speaker is more frequently prone to use borrowed function words to better explain the connections between different lexemes in a sentence or between different sentences.

3.1.1. Adapted borrowings

Nominal borrowings are often integrally acquired, but a number of occurrences of adapted loans are also recorded in the corpus (9):

- (9) a. *albergos* ‘hotels’, *interesses* ‘interests’, *piscinas* ‘swimming pools’, *genitores* ‘parents’, *cuginus* ‘cousins’, *ninnanannas* ‘lullabies’.
b. *filmis* ‘movies’.
c. *golfos* ‘jumpers’, *otellos* ‘hotels’.
d. *iscelta* ‘choice’, *ferragostu* ‘mid-August’, *nonnu* ‘grandfather’,¹⁵ *afiatamentu* ‘fellowship’.

The adaptation is typically morphological in nature (it involves plural inflections (9a)), but it can also concern phonological choices (9d): the item *iscelta*, for example, yields a prosthetic vowel before an initial /ʃ/, a phonological process which can be found mainly in Central and Northern Sardinian varieties.¹⁶ As we see in (9b), however, it is also possible to find both phonological and morphological adaptations: in *filmis* ‘movies’, an epenthetic vowel is added to the word ending in consonant (as is also the case of Italian spoken in Sardinia) and the plural is formed

with the addition of the morpheme *-s/* to the regularised form: thus, from *filmi* ‘movie’, we have the regularised plural *filmis*. In (9c) the items *otellos* ‘hotels’ and *golfos* ‘jumpers’ are noteworthy. Probably due to the high frequency of words ending in consonants in French, these forms are specific to the *FRA_Sard* corpus. Both are based on a word ending in consonant; however, while the former is clearly based on French *hotel*, the latter derives from Italian *golf* (itself a borrowing from English *golf*). Plural forms in *-os* may presuppose a singular form in *-u* (which is the most regular Sardinian morpheme of masculine nouns), or more rarely, in *-o* (e.g. Nuoro *coro*, *-os* ‘heart’), and are masculine as in the model language.

As for verbs, the typological closeness between Sardinian and Italian allows the same flexible class of the model language to be maintained, as in (10):

- (10) *garantit* ‘ensure.3SG’, *sofocavat* ‘smother.PST.3SG’, *pigliat* ‘take.3SG’, *rinfaciat* ‘taunt.3SG’, *coltivae* ‘grow.INF’, *si svolgit* ‘take_place.3SG’.

As far as adverbs in *-mente* are concerned, we must consider that the forms with a final *-i* show the typical phonetical adaptation of Southern Sardinian, which displays a general raising of mid- final vowels *-e* and *-o* respectively to *-i* and *-u*:¹⁷

- (11) *naturalmenti* ‘naturally’, *sinceramenti* ‘frankly’, *regolarmenti* ‘regularly’, *giustamenti* ‘rightly’, *normalmenti* ‘normally’, *immancabilmenti* ‘unfailingly’.

In the adjectives we observe both phonological and morphological adaptations: according to Sardinian morphological system, we observe forms with *-u* for the singular masculine (*-o* is the corresponding morpheme in Italian) and we can also see the morpheme *-s/* which marks masculine and feminine plurals.

- (12) *picolu* ‘little.M.SG’, *picolas* ‘little.F.PL’, *allaciadas* ‘fastened.F.PL’, *consistentes* ‘considerable.PL’, *estraneas* ‘extraneous.F.PL’, *sigillada* ‘sealed.F.SG’, *coinvoltu* ‘involved.M.SG’.

As far as phonology is concerned, it is useful to note that the feminine past participle of the first conjugation of Italian *-ata* is regularised with the final sequence *-ada*¹⁸ in the written corpus.

Finally, the peculiar cases of the preposition *duranti* ‘during’ (cf. Italian *durante*) and the conjunction *dopu* ‘after’ (cf. Italian *dopo*) are worth noting. As we saw above, the raising of a mid-final vowel (cf. *-e* > *-i* and *-o* > *-u*) is clearly connected with Southern Sardinian varieties. Indeed, other phonological adaptations cannot be detected in our written corpus. As a matter of fact, in Southern Sardinian varieties,

metaphonic alternations are more complex than Central and Northern ones, as metaphony is only triggered by the final high vowels *-u* and *-i*, but not by final high vowels resulting from a word final raising.¹⁹ This is true not only for the hereditary lexicon but also for many loans.²⁰ For this reason, in the original audio of the interviews, the Southern item *dopu* is always pronounced [dɔ̃p^u]. Such pronunciation reveals the adaptation of an intervocalic voiceless stop of the model language; as Contini (1987) and other scholars have claimed (cf. Virdis 1978: 50; Bolognesi 2012: 161; Lai 2022a: 609) Sardinian varieties usually do not display contrastive consonantal length, and Sardinian simple consonants are usually longer than the Italian ones. For this reason, as Bolognesi (2012: 161) points out, Sardinian people are usually not able to “produce contrastive consonantal length when speaking Italian” as well.

3.1.2 *Multiword contact phenomena*

Among lexical contact phenomena, corpus evidence proves to be particularly fruitful in the area of phraseology. In particular, collocations and multiword units deserve special attention. Both the Italian and the French subcorpora are characterised by the presence of multiword borrowings.

| | ITA_SARD CORPUS | FRA_SARD CORPUS |
|--|-----------------|-----------------|
| Multiword nouns | 37 | 17 |
| Multiword verbs | 5 | – |
| Multiword adjectives | 6 | 3 |
| Multiword adverbs | 35 | 12 |
| Multiword prepositions | 6 | 1 |
| Multiword conjunctions | 3 | 1 |
| Multiword pronouns | 5 | 1 |
| Interjections, formulas, discourse markers | 8 | 4 |

Table 2. Set of multiword borrowings identified in the corpus.

They typically belong to specialised areas of the lexicon and highlight the lack of the specific terminology in the replica language. Nominal multiwords are the greatest set, followed by the one of adverbial units. With the exception of the combination *medaglias al merito* ‘medals of merit’, which shows the Sardinian nominal inflection of the head of the NP, nominal multiwords are typically non-adapted word combinations, as in (13):

- (13) *gioco di squadra* ‘teamwork’, *parità dei diritti* ‘equality of rights’, *generi alimentari* ‘foodstuffs’, *estensione geografica* ‘geographical extension’.

Nominal multiwords may also emerge as the result of contact of both Italian and French into Sardinian: this is the case of *assistente mater-nelle* ‘maternal assistant’, from the French subcorpus, where the first element is kept from Italian and the second one from French. Nominal multiword borrowings also include non-lexicalised sequences, as the ones represented in (14-16):

- (14) *una coppia de (antzianos)* ‘a couple of (elderly people)’, *una cerchia de (amigos)* ‘a group of (friends)’.
(15) *una specie de (sede sociale)* ‘a sort of (head office)’.
(16) *una cosuccia de (nudda)* ‘a little thing of (nothing)’, *una roba de (su gènere)* ‘a thing of (this type)’.

All the sequences are clear instantiations of the so-called ‘light nouns’ (cf. Simone & Masini 2014), namely nominal elements which have undergone a reduction of referential force and have acquired a grammatical function. Borrowed light nouns typically occur as syntactic head of the [Noun₁ Prep Noun₂] semi-lexicalised pattern (i.e. Noun₁), and are used to (i) provide a quantitative information about Noun₂, as the quantifiers in (14); (ii) “modulate the extension of Noun₂ by weakening its belonging to a definite category” (Simone & Masini 2014: 57), as the approximators in (15); (iii) introduce the Noun₂ as a generic reference or an encapsulation device, as in (16), which are partial calques from Italian, and fall under the class of ‘solid’ multiword nouns²¹ in the terms of Simone 2006). Interestingly, in the sequences in (14-16), the light noun (i.e. Noun₁) is integrally borrowed from Italian, whereas Noun₂ is often of Sardinian origins. These uses can be considered the result of code-mixing phenomena with a phrasal scope: the phrasal head is loaned, while the modifier is Sardinian. It is also worth noting that, in the case of quantifiers (14), the informants also employ Sardinian light nouns, depending on the context (e.g. *una pariga de* ‘a couple of’, *unu muntone de* ‘a lot of’, *unu bucone de* ‘a bite of’). Furthermore, light nouns having the role of classifiers (namely those signalling the semantic class of Noun₂), only appear in the Sardinian original form in the corpus – e.g. *una pisedda de (casu)* ‘a wheel of (cheese)’, *una soma de (trigu)* ‘a measure of (wheat)’ –: this may be due to the fact that nouns in the Noun₁ position show a sort of semantic solidarity with Noun₂, thus the whole phrase is more likely to be perceived as a cohesive and lexicalised unit.²²

Interestingly, differently from nominal units, verbal collocations or verbal multiword sequences are often adapted to Sardinian inflectional system or mixed to Sardinian lexemes, as the following examples show:²³

- (17) a. *ITA_Sard*: F-1940 Ovodda (Bergamo)
 s' *erricu* prendet s' aereo
 the.SG rich.M.SG take.3SG the.SG plane
 'rich people take the plane'
- b. *ITA_Sard*: M-1947 Laerru (Gallarate)
muzere *mia* no aiat polso
 wife my.F.SG not have.PST.3SG wrist
 'my wife didn't show character'
- c. *ITA_Sard*: F-1966 Villaverde (La Spezia)
non mi *ndi* *podia* *mancu* rendi contu
 not me of it can.PST.1SG not_even render.INF account
 'I couldn't even realise it'
- d. *ITA_Sard*: F-1940 Ovodda (Bergamo)
perciò *bisonzavat* *a* si *rimboccare* *sas* *manos*
 so be_necessary.PST.3SG to REFL tuck.INF the.F.PL hand.PL
 'so sleeves needed to be rolled up'
- e. *ITA_Sard*: M-1943 Nuoro (Brescia)
a su nessi *ressesit* *a* tirar *a foras*
 at_least manage.3SG to pull to out
 '[she has a passion for reading] so she can get [all her worries] out'

On the one hand, adaptation may concern only the verb and its inflections (17a-b), or both the verb and the noun (17c). On the other hand, a part of the multiword unit may be borrowed integrally: the noun (as *aereo* 'plane' or *polso* 'wrist', in (17a-b)) or the verb (as *rimboccare* 'tuck' or *tirare* 'pull' in (17d-e)). Multiword verbal loans are extremely rare in the corpus, with respect to other phraseological units. Furthermore, verbal multiword borrowings never employ specialised terms. This could suggest that the need for multiword borrowings is mainly due to onomasiological reasons.

In the corpus, multiword adverbs and adjectives are frequent (the former more than the latter).²⁴ Adverbial units are more frequent than adjectival ones in the corpus. They are fully lexicalised sequences which show a high degree of cohesion (Simone 2007) and an idiomatic semantics (Casadei 1995, 1996) or a pragmatic value (Simone 2007). A part few exceptions, they are typically not adapted (e.g. *pianu pianu*), since they are generally used as integral borrowings. From the structural point of view, there is a slight preference for the configuration

[Prep + (Det) + Noun] (18a-b), which is represented in 45% of multiword adverbials. However, also other patterns are represented (cf. 18c-g).²⁵

- (18) a. *per forza* ‘necessarily’, *a fondo* ‘thoroughly’, *tra virgolette* ‘in quotes’, *di solito* ‘usually’.
b. *alla fine* ‘at the end’, *al massimo* ‘at most’.
c. *man mano* ‘little by little’, *piano piano (pianu pianu)* ‘little by little’, *quasi quasi* ‘almost’.
d. *ogni tanto* ‘sometimes’.
e. *a mano a mano* ‘little by little’, *in fondo in fondo* ‘after all’, *di tanto in tanto* ‘occasionally’.
f. *più o meno* ‘more or less’.
g. *su per giù* ‘more or less’.

From the semantic standpoint, many examples have a temporal meaning (e.g. *di tanto in tanto* ‘occasionally’) or are employed as pragmatic devices (e.g. *su per giù* ‘more or less’). An important feature concerning adverbial units is related to the level of ‘constructional schemata’²⁶ employed to build indigenous sequences. Indeed, language contact may not only involve fully lexically specified word combinations (cf. at least Fillmore *et al.* 1988), but also “abstract structural patterns in the two languages that are functionally and formally similar but not identical” (Pietsch 2008: 215), thus, developing a sort of cognitive “interlingual identification” (Weinreich 1953). This clearly presupposes the availability of formal correspondences between the two languages (Pietsch 2008). In particular, a convergence between adverbial constructional schemata concerns those units characterised by the presence of the preposition *a* ‘to/at’ followed by a noun (i.e. the pattern [*a* + Noun]). On the one hand, this pattern characterises Sardinian indigenous multiword adverbs (e.g. *a piticu*, lit. ‘at short, slowly’; *a pustis* ‘after’), and on the other hand, the same configuration is employed in the creation of calques from Italian (e.g. *a pena* ‘almost’, *a posta* ‘intentionally’). Interestingly, the pattern is used as a sort of structural calque from Italian: indeed, the same structure is used in Sardinian even to replicate Italian multiword sequences which do not employ the preposition *a*: this is the case of Italian *da sola* and the Sardinian *a sola* ‘alone’).

Only few instances of adjectival units are recorded in the corpus, as those in (19):

- (19) *a fondo chiuso* ‘close-hand’, *a tre stelle* ‘three-star’, *in dotazione* ‘at disposal’, *vero e proprio* ‘real’ (lit. ‘true and proper’).

As happens with adverbial units, also multiword adjectives are structurally heterogeneous: despite the low number of examples, four different patterns are represented in (19).²⁷ However, their occurrences are not remarkable, neither in terms of structures, nor in terms of fre-

quency of use. The only remarkable feature is that they could be necessity loans, as happens for adverbial units.

In the corpus, among multiword units it is possible to find few examples of complex prepositions, as the examples in (20):

- (20) *rispetto a* ‘with respect to’, *oltre a* ‘beyond’, *contro a* ‘against’, *in confronto de* ‘in comparison with’.

Similarly to multiword verbal expressions, also this group of combinatorial units is not rich; this is mainly due to the fact that, as happens for verbal sequences, borrowed prepositional units co-exist with indigenous ones, which frequently occur in the corpus (e.g. *a foras de* ‘out of’, *a curtzu de* ‘close to (lit. ‘at short of’)', *a medade de* ‘in the middle of’, *a diferèntzia de* ‘differently from’).

Corpus analysis has also highlighted groups of combinatorial sequences having a pragmatic function. Among the others, interjections (21a), formula (21b) and discourse markers (21c) (cf. §3.3) deserve special attention:

- (21) a. *guai a chi me la tocca!* ‘woe betide anyone who touches it!’, *porca di una miseria!* ‘holy cow!’, *beato porco!* ‘lucky pig!’, *cavolo* ‘oh boy!’.
b. *figlia mia* ‘my beloved daughter’, *come stai?* ‘how are you?’.
c. *figurati* ‘please’, *e basta* ‘and that’s it’, *tant’è vero* ‘so true’.

3.2. Contact phenomena from the syntactic viewpoint

The use of linking words and pronominal structures introducing subordinate sentences is one area where superstrate languages – in this case Italian and French – seem to have a significant impact on the language of informants in our corpus. In particular, connectives or pronominal structures that introduce subordinates are worth noting.

Even in the French subcorpus, several Italian linking words are systematically used without being adapted to the phonetic-phonological system of Sardinian.²⁸ The following examples are taken from the Italian and French subcorpora:

- (22) *ITA_Sard*: M-1947 Laerru (Gallarate)
an iscopertu s’ àtera e quindi devio seberare!
have.3PL discover.PTCP.M.SG the.SG other.F.SG and therefore must.PST.1SG choose.INF
‘they discovered the other and so I had to choose!’

- (23) *ITA_Sard*: F-1965 Sarule (Roma)
però non so andà a chircare sos sardos
but not am gone.PTCP.F.SG to look_for.INF the.M.PL Sardinian.M.PL
‘but I didn’t go looking for Sardinian people’

(24) *ITA_Sard*: F-1966 Villaverde (La Spezia)

iap' ai fatu sa fini de mamma mia cioè fillus mius
 COND.1SG have.INF done.M.SG the.F.SG end of mother my.F.SG that_is son.PL my.M.PL
si nd' iant essi andaus
 REFL from_here COND.3PL be.INF go.PTCP.M.PL
 'I would have done what my mother did... that is my children would have left!'

(25) *FRA_Sard*: M-1935 Borore (Le Creusot)

nos vidimus pius in sos interros nos saludamus
 we see.1PL more in the.M.PL funeral.PL REFL greet.1PL
chin totus mais ici totu
 with all.M.PL but here all.M.SG
 'we mostly meet (lit. see us) at funerals, but no more than that'

A similar situation characterises the pronominal structures with which speakers often introduce subordinate propositions. Relative sentences deserve special attention: as a matter of fact, the inflected forms of relative pronouns (e.g. Italian *il quale / la quale* 'which' or *cui* 'to whom'), are totally absent in Sardinian (cf. DES: 205; Putzu 2011);²⁹ indeed, they are usually replaced by the indeclinable *chi* 'who/which' with the repetition of the clitic, as in the following examples:

(26)

a. Nuoro

su pittsinnu ki l áppo dáu zu vuzîle
 the.M.SG boy REL to_him have.1SG give.PTCP.M.SG the.M.SG rifle
 'the guy I gave the rifle to'

b. Pula

sa ßittfókka yi anqámuz a iskóla umpári
 the.F.SG girl REL go.PST.1PL to school together
 'the girl we went to school with'³⁰

As we see in (27) this strategy regularly occurs in our corpus:

(27) *ITA_Sard*: M-1925 Siniscola (Fiumicino)

tando b' it su presidente chi nd' it [...] ma sicomente it
 so there be.PST.3SG the.M.SG president REL there be.PST.3SG but since be.PST.3SG
una persona chi li piaghiat meta a organizzare
 a.F.SG person REL him like.PST.3SG a_lot to organise.INF
 'then there was the president who was [...], but as he was a person who really liked to organise a lot'

The necessity of pronominal borrowing is often imposed by the topics and, perhaps, by the partially artificial context of the semi-directorial interview. This leads the informants to transfer the Italian pronominal structures to Sardinian sentences, as in the following examples:

- (28) a. *ITA_Sard*: F-1948 Martis (Roma)
non bi sun limbas in cui no si podet nàrrere calesiat cosa
 not there be.3PL languages in REL not IMPERSONAL can.3SG say.INF any thing
 ‘there is no language in which everything cannot be said’
- b. *ITA_Sard*: M-1968 Samugheo (Rivoli)
per cui geo andao in Sardegna
 for REL I go.PST.1SG in Sardinia
 ‘so, I went to Sardinia’
- c. *FRA_Sard*: F-1935 Nuragus (Lyon)
deu no bollu a èssiri ind una idda in cui depu fai a gestus
 I not want.1SG to be.INF in a.F.SG place in where must.1SG do.INF at gesture.PL
 ‘I don’t want to be in a country where I have to make hand gestures’

Sentences in (28) deserve special attention. In (28a), the speaker prefers the Italian structure *in cui* ‘in which’ to the Sardinian relative *chi*, although she chooses the Sardinian connective *gasichì* ‘so that’ (probably a calque of Italian *cosicché*). This choice may be coherent with the ‘high’ register of the topic addressed. In (28b), Italian *per cui* does not really have the value of a relative pronoun; instead, it seems to be used as a connective, as in the model language. Finally, it is worth analysing the pronominal structure *in cui* ‘in which’ of the example (28c). The use of this structure may have been preferred due to the homophony between Italian *cui* ‘(of/to) which’ and Central-Southern Sardinian *cui* ‘where’. Therefore, in the latter case, the influence of the model language (i.e. Italian) is less direct but cannot be excluded *a priori*. Nevertheless, in some cases, the relative pronoun may be inflected for gender and number as *su cale* / *sa cale* ‘of which’ (which follow Italian *il quale* / *la quale* ‘which’), that are also widely used by the speakers in our corpora), as in the examples in (29):

- (29) a. *ITA_Sard*: M-1943 Ittiri (Roma)
un’ àtera cosa importante pro sa cale nos devimus battere
 a.SG other thing important for the.F.SG REL REFL must.1PL fight.INF
e impignare est sa continuidade territoriale
 and strive.INF is the.F.SG continuity territorial
 ‘another important thing we have to fight and strive for is the territorial continuity’
- b. *ITA_Sard*: M-1948 Siligo (Bergamo)
unu mundu in su cale est resessida a ligare
 a.M.SG world in the.M.SG REL is manage.PTCP.F.SG to bind.INF
 ‘a world in which she was able to forge friendships’

As far as the information structure is concerned, an interesting phenomenon comes from the French subcorpus. The use of the cleft sentence is extremely frequent among Sardinian speakers living in France (it occurs 20 times in the whole French subcorpus). In the following, some examples are provided:

(30)

- a. *FRA_Sard*: F-1935 Nuragus (Lyon)
est aci chi apu cannotu sorri tua
 is how REL have.1SG know.PTCP.M.SG sister your.F.SG
 ‘that’s how I met your sister’
- b. *FRA_Sard*: F-1948 Orani (Le Creusot)
su bellu chi acato de ciambau est chi sa cultura est accessibile
 the.M.SG nice.M.SG REL find.1SG of change is that the.F.SG culture is accessible
 ‘what I find amazing is that culture is accessible [to...]’
- c. *FRA_Sard*: F-1935 Nuragus (Lyon)
est issa chi depit progetai sa sicuresa nel lavoro
 is she REL must.3SG design.INF the.F.SG safety in.the.M.SG work
 ‘she is the one who has to design safety at work’
- d. *FRA_Sard*: F-1935 Nuragus (Lyon)
mi parit chi est sa chi at presentau mellus is sardus
 to_me seems that is the_one.F.SG REL has present.PTCP.M.SG better the.PL Sardinian.PL
 ‘it seems to me that she is the one who presented the Sardinians best’
- e. *FRA_Sard*: F-1935 Nuragus (Lyon)
est issu chi s’ at riconnotu
 is he REL us has recognise.PTCP.M.SG
 ‘he is the one who recognised us [as a Sardinian Association]’
- f. *FRA_Sard*: M-1935 Borore (Le Creusot)
est sa fèmina chi dat sa vida
 is the.F.SG woman REL give.3SG the.F.SG life
 ‘it is the woman who gives life!’

It is worth noting that this phenomenon characterises all informants taking part in the interviews transcribed. As is well known, the cleft sentence is mainly used in French “mainly because it is the only way to overcome the rigid word order of this language” (D’Achille *et al.* 2005: 250, translation ours).³¹ If the use of the cleft sentence is occasional and almost irrelevant in the case of the speakers of the Italian subcorpus, the number of this kind of sentences grows considerably in the interviews carried out in France.

3.3. Contact phenomena from the pragmatic viewpoint

The area of discourse markers is particularly permeable to contact phenomena (Dal Negro & Fiorentini 2014). Discourse markers are linguistic elements which are devoid of referential value and are very commonly used in speech. It has been shown that in the multilingual contexts involving Italian and other local varieties, speakers frequently employ discourse markers of the model language (Fiorentini 2017). In particular, several studies have shown that discourse markers can strongly facilitate interaction in multilingual contexts. In the context of this paper, it is not possible to discuss neither the various classifications of these pragmatic elements proposed in the literature (cf. Fischer 2006; for Italian, cf. Bazzanella 2006) nor the identification of their various functions. Since the corpus analysis allows us to identify the multiple functional correlates of the use of code mixing, we therefore limit to organising the units identified in the corpus according to their function in context. In particular, the following functional classes are taken into account for the purpose of this analysis: (i) interactional markers, (ii) metatextual markers, (iii) modal markers, (iv) reformulation markers, (v) general extenders.

As far as INTERACTIONAL MARKERS are concerned, they can be used to anchor the speaker's utterance in the turns of the speech architecture. The speaker uses these elements in the initial or final position of her/his turn, when she/he wants to signal that she/he is about to take or leave the floor to someone else. They occur in both the French (31) and the Italian (32) subcorpora:

(31) *FRA_Sard*: M-1935 Borore (Le Creusot)
A torrare goi in vacantzà ma po bìvere nono! E voilà!
to come_back.INF in_this_way in holiday but for live.INF no and here_is
'Going back to Sardinia on holiday could be fine, but to live there...no! Voilà!'

(32) *ITA_Sard*: M-1949 Vilamar (Livorno)
e vabbè cantus annus funt immoi?
and ok how_many.M.PL year.PL be.3PL now
'and ok, how long?'

The use of these discourse markers can also be justified by the speakers' willingness to plan the discourse (Fedriani & Sansò 2017).

METATEXTUAL MARKERS are used to structure complex textual units. These elements link different sentences, highlighting logical and argumentative connections between parts of discourse, to produce coherence. Thus, they are fundamental in the planning and organisation of the discourse (Halliday & Hassan 1989). Corpus analysis reveals that they occur in both the Italian and French subcorpora:

Notes and updates on language contact between Sardinian and Italian/French

- (33) *ITA_Sard*: F-1940 Ovodda (Bergamo)
dio 'ambiare *medas* 'osas, però purtroppo *non si* *podet*
 COND.1SG change.INF a.lot.F.PL thing.PL but unfortunately not IMPERSONAL can.3SG
 'I should change many things but unfortunately one cannot do it!'
- (34) *FRA_Sard*: F-1935 Nuragus (Lyon)
e insomma in Frància est nàsciu su de Brenne
 and in_short in France is born.PTCP.M.SG the_one.M.SG of Brenne
 'and... in short, in France the one [Sardinian club] of Brenne was born first'

Several examples of MODAL MARKERS can be identified in the corpus. They are used to signal the modal value of the utterance. These may have evidential value, signalling that (35) the content of the sentence has not been verified by the speaker, or that (36) the speaker explicitly expresses his commitment to the truth of the utterance:

- (35) *ITA_Sard*: F-1940 Ovodda (Bergamo)
deo no nd' isso ite est 'i sutzedet in drinto de sa conca de
 I not of.it know.1SG what is REL happen.3SG in inside of the head of
sos 'i nos governan, però credete-mi, badiade bos bene in s'isprigu,
 the.one.M.PL REL us govern.1PL but believe.IMP.2PL-me look.IMP.2PL you well in the mirror
rifletide poite sa zente sich' andat [...]
 reflect.IMP.2PL because the.F.SG people REFL.from here go.3SG
 'I don't understand what's going on in the heads of those who govern us, but believe me, take a look in the mirror and reflect because the people are leaving their own country!'
- (36) *ITA_Sard*: M-1943 Nuoro (Brescia)
per essere sincero sa prima borta chi so partiu in su sessantanobe
 for be.INF honest the.F.SG first time REL am leave.PTCP.M.SG in the.M.SG sixty_nine
semus ghiraos dopo bindichi annos
 be.1PL come_back.PTCP.M.PL after fifteen year.PL
 'to be honest, the first time I left, in 1969, we came back after 15 years'

Among the Italian borrowings, REFORMULATION MARKERS are worth noting. They are used to signal the presence of a reformulation or an exemplification (37-38). Also reformulation markers appear in both the Italian and the French subcorpora:

- (37) *ITA_Sard*: F-1955 Neoneli (Vercelli)
nebodes mios, per esempio, sos de Neoneli ant
 nephew.PL my.M.PL for example the_one.M.PL of Neoneli have.3PL
sighiu a dda coltivate
 follow.PTCP.M.SG to her grow.INF
 'my nephews, for example, from Neoneli are still practicing it'
- (38) *ITA_Sard*: F-1940 Ovodda (Bergamo)
nois 'omente emigraos podimus batire in Sardigna
 we as emigrant.M.PL can.1PL bring.INF in Sardinia

s'esperientza de vida 'i tenimus ino'e, ad esempio,
 the.SG experience of life REL hold.1PL here for example
nois semus in Lombardia ma sos sardos sunis in totu su mundu!
 we be.1PL in Lombardia but the.M.PL Sardinian.PL be.3PL in all the.M.SG world
 'we as emigrants can bring to Sardinia our experience we gained here *for example*: we are
 in Lombardia but Sardinians are all over the world!'

Finally, the class of general extenders can be noted. General extenders are linguistic elements devoid of any referential value, which are used to indicate the presence of a potential extension of an utterance, which is not made explicit. From the structural point of view, the general extenders which have been identified in the corpus tend to have the structure of coordinated phrases employing the conjunction *e* 'and', as the following examples show:

(39) *ITA_Sard*: M-1943 Serramanna (Alessandria)
invècias is atrus mancai pagant afitus e compagnia bella
 instead the.PL other.M.PL maybe pay.1PL rent.M.PL and company nice.F.SG
 'whereas the others have to pay a rent *and so on*'

(40) *ITA_Sard*: M-1947 Laerru (Gallarate)
mi devio samunare e via dicendo
 me must.1SG wash.INF and way saying
 'I needed to wash up *and so on*'

4. Conclusions

As it is common cross-linguistically, contact phenomena mainly affect the lexical level. Against our expectations, however, most borrowings concern very common and frequent lexical entities (i.e. the ranges or fundamental and common vocabulary, cf. De Mauro 1999b), and only to a lesser extent specialised terms. Furthermore, the number of lexical loans is higher than the number of syntactic or pragmatic borrowings. Nevertheless, the frequency of occurrence in our corpus varies significantly between the different classes of borrowings: while lexical contact phenomena are low in frequency, the occurrence of pragmatic and syntactic ones is much more consistent in the corpus.

From the syntactic point of view, the higher degree of hypotaxis of Italian and French is transferred to Sardinian. In particular, this is evident in the use of subordinating conjunctions, relative pronominal structures and cleft sentences, thus confirming the general tendency of languages that are mainly spoken to introduce exogenous subordination strategies, when exposed to the contact with a more elaborated (model)

language (Chafe 1985, Mithun 2012). Therefore, more elaborated textual syntactic structures emerge via language contact.

From the pragmatic viewpoint, corpus data highlight a major use of exogenous discourse markers by the informants. This issue confirms well-known tendencies identified in the literature concerning the use of discourse particles in contact varieties.

This study has shown that the field of heritage Sardinian, spoken in Italian contexts (or in those of other major Romance varieties such as French), is a fertile domain for the inquiry of language contact, and still deserves further attention. Also, this investigation has demonstrated that collection of spoken data and the use of corpus linguistics methods may highlight new aspects of contact phenomena, thus providing an alternative perspective for additional generalisations.

Abbreviations

1, 2, 3 = first, second, third person; COND = conditional; F = feminine; IMP = imperative; INF = infinitive; M = masculine; PL = plural; PST = past tense; PTCP = participle; REFL = reflexive; REL = relative pronoun; SG = singular.

Acknowledgements

We are particularly grateful to the editors of this volume and to the anonymous reviewers for their comments and suggestions. Although the paper is the fruit of a joint reflection and cooperation, for academic purposes Simone Pisano is responsible for sections 1, 2, 2.1, 3.1.1, 3.2; Valentina Piunno for sections 3.1, 3.1.2; Vittorio Ganfi for section 3.3. Section 4 is to be attributed to all authors.

Notes

¹ Cf., at least, Thomason & Kaufman (1988), Aikhenvald & Dixon (2001), Thomason (2001), Winford (2003).

² Cf. Hickey (2010) for an in-depth overview.

³ It should be noted, however, that Barbato (2017: 158-159) devotes a whole paragraph to strictly grammatical issues on the Catalan *superstratum*.

⁴ On this issue, cf. the pioneer work by Dorian (1973: 413-438).

⁵ On this topic, cf. also Kupisch (2021: 46-49) and Lai (2022b: 40-41).

⁶ The project has been funded by the Autonomous Region of Sardinia. More specifically, its realisation was made possible thanks to the contribution of the *Assessoradu de s'istrutzione pùblica, benes culturales, informatzione, ispetàculu e isport*, of the

Assessoradu de su traballu, formatzione e seguràtzia sotziale, of the involvement of the network of the Federation of Sardinian Associations in Italy and, to a lesser extent, of the Federation of Sardinian Associations in France.

⁷ The French Sardinian data were collected from speakers living in Le Creusot and in the Île de France; the Italian Sardinian data stem from informants living in Milano, Brescia, Bergamo, Udine, Torino, Alessandria, Vercelli, Genova, La Spezia, Padova, Vicenza, Bologna, Parma, Firenze, Livorno, Pisa, Siena, and Roma.

⁸ The informants of the non-Sardinian varieties are 15; the remaining interviews involve speakers of Sardinian varieties.

⁹ This study equally considers phenomena related to the influence of French and Italian. As noted by one of the two anonymous reviewers, dealing with these two different conditions at the same time may seem rather odd (i.e. Sardinian spoken in the Italian peninsula and Sardinian spoken in France), since in the case of the 'Italian Sardinian' the dominant language is represented by the national and official language of the State (on this topic, cf. Montrul 2016: 15). Nevertheless, we believe that the two scenarios are comparable in terms of language contact and influence of the model language. In the case of the 'French Sardinian', we interviewed people with medium-low education who migrated very early and received very limited influence by the Italian language.

¹⁰ Since the earliest studies on heritage languages, the intergenerational patterns of language shift from heritage to dominant languages have been emphasised (Silva-Corvalán 1994). For these reasons, in the scholarly debate concerning heritage languages, informants are subdivided according to the generational distance from their arrival (first and second generation speakers). Nonetheless, it is worth noting that all the informants whose interviews were transcribed are first generation speakers. Furthermore, most informants are in their fifties or more, given the high median age in the Associations they belong to. Nevertheless, the transcribed material is sufficiently balanced against the speakers' age, since it also includes some informants in the age-range 30-50.

¹¹ That is, the only language they acquired in their childhood was Sardinian. All informants are L2 speakers of Italian, since they learnt Italian in primary school age.

¹² It would be extremely interesting to compare contact phenomena among Gallurese, Sassarese, Tabarchino and Catalan of Alghero, and to analyse the differences between them and the Sardinian varieties. Nevertheless, the data collected in the corpus are not balanced against the representativeness of all languages; in fact, the number of interviewees speaking non-Sardinian varieties is much lower than the other group. Therefore, in this paper we prefer to focus only on Sardinian varieties.

¹³ This operation had some limitations, due to the lack of a Sardinian dictionary compatible with the software used, through which the automatic annotator could recognise and label words, thus associating each graphic form to the labels of the appropriate word classes and lemmas. However, the software provided very good results concerning the identification of word boundaries and sentence boundaries.

¹⁴ See the frequency-based distinction in the Italian vocabulary proposed by GRADIT (De Mauro 1999a, 1999b). The basic vocabulary includes the 'fundamental vocabulary', the 'high usage vocabulary', and the 'high availability vocabulary'; it contrasts with the common vocabulary, i.e. "words belonging to different disciplines and areas generally known to people having secondary education" (Chiari & De Mauro 2012: 27).

¹⁵ In Sardinian, *nonnu* often means 'godfather', while the term for 'grandfather' may vary geographically (*jaju*; *manneddu*; *mannoi* etc. cf. DES: s.v. *mánnu*).

¹⁶ In Central and Northern Sardinian, the prosthetic vowel /i/ always appears before /s/ + consonant clusters.

¹⁷ As shown below, these final *-i* < *-e* and *-u* < *-o* do not trigger metaphony; hence,

as shown in the original video-interviews, our Southern Sardinian native speakers always realise [-mɛnti]. However, in loanwords this phenomenon is not systematic. For an exhaustive discussion on this topic, cf. Viridis (1978: 26), Bolognesi (2012: 19-22), Krefeld (2017: 330-331), Molinu (2017: 350-352) and Lai (2022a: 600-601).

¹⁸ In this case the grapheme < d > is used for a [ð] sound.

¹⁹ See, for instance, the following examples: Pula [bɛntu] < VENTU ‘wind’; Pula [óllu] < OLEU ‘oil’, but Pula [dɛʒi] < DECE ‘ten’; Pula [ʒttu] < OCTO ‘eight’.

²⁰ Cf. Pula [televizióni] ‘television’ (cf. it. *televisione* realised as [televizjónɛ] in Sardinian Italian). One must keep in mind, however, as claimed by Lai (2022a: 612), that in loanwords “metaphony is not systematic”.

²¹ Namely, nouns which are semantically light and are used in the [Noun₁ Prep Noun₂] pattern to establish a reference (Simone 2006).

²² It is worth noticing that an ethnographic reason may also contribute to the preference for indigenous light noun classifiers, namely, the relationship between code and typical context of use.

²³ Hereinafter, when the context of occurrence is given, the examples will contain the following information: subcorpus (*ITA_Sard/FRA_Sard*), sex (male/female), age (year of birth), place of birth and place of current residence. When the examples are given in their context of use, the lexemes under discussion (e.g. loans) are deitalicised.

²⁴ Cf. Piunno (2018) for a general description of the multiword adverbial and adjectival units in Italian, with reference to those with a PP structure.

²⁵ The different units have been separated according to their structural patterns, as follows: [Prep Noun], [Prep Det Noun], [Noun₁ Noun₂], [Adj Noun], [Prep Noun Prep Noun], [Adv Conj Adv], [Adv Prep Adv].

²⁶ This concept is taken from the theoretical framework of Construction Grammar (cf. Goldberg 1995, among others). ‘Constructional schemata’ refer to the abstract “cognitive schema, that is, a mental representation that captures the construction’s general traits” (Hilpert 2014: 5). As Hilpert notes, “many idioms cannot be stored as fixed strings, which makes it necessary to think of idiomatic expressions as schemas with slots that can be filled with certain elements but not others” (2014: 6).

²⁷ The patterns are [Prep Noun], [Prep Noun Adj], [Prep Adj Noun], [Noun₁ Conj Noun₂].

²⁸ Note that the intervocalic voiceless stops usually display lenition both in Northern and Southern Sardinian varieties (but not in the centre of Sardinia). For this feature see the following example (taken from spontaneous speech recordings, fieldwork conducted in Pula in 2018):

| | | | | | |
|--------|---------------------|-----------|-----|------|----------------------|
| (Pula) | déu | ɣúndi | nɔ | dɔdu | ʃʃiém ^m u |
| | I | therefore | not | that | knew |
| | ‘So I didn’t know!’ | | | | |

As we can see in the item *ɣúndi* we clearly find some adaptations to the phonological system of Southern Sardinian varieties which display the lenition of voiceless stops. Regarding the Italian sequence [kwin-], it is realised as two separate syllables.

²⁹ The adapted inflected forms *su cale / sa cale* ‘which’, documented since the Middle Ages, are probably due to an early influence of medieval Tuscan. As a matter of fact, they seem quite unnatural in spoken Sardinian (cf. Pisano 2017).

³⁰ These two examples are taken from semi-spontaneous speech recordings and are the result of fieldwork conducted in Nuoro and Pula in 2018.

³¹ For further details about cleft sentences in Italian, cf. Benincà (1978), Berruto (1987), Benincà *et al.* (1988), and Berretta (1994, 2002).

Bibliographical References

- Aikhenvald, Alexandra Y. & Dixon, Robert M. W. 2001. *Areal Diffusion and Genetic Inheritance: Problems in Comparative Linguistics*. Oxford: Oxford University Press.
- Aikhenvald, Alexandra Y. 2006. Grammars in contact: A cross-linguistic perspective. In Aikhenvald, Alexandra Y. & Dixon, Robert M. W. (eds.), *Grammars in Contact: A Cross-Linguistic Typology*. Oxford: Oxford University Press. 1-66.
- Barbato, Marcello 2017. Superstrato catalano. In Blasco Ferrer *et al.* (eds.) 2017. 150-167.
- Bazzanella, Carla 2006. Discourse markers in Italian: Towards a 'compositional' meaning. In Fischer, Kerstin (ed.), *Approaches to discourse particles*. Amsterdam: Elsevier. 449-464.
- Benincà, Paola 1978. Sono tre ore che ti aspetto. *Rivista di Grammatica Generativa* 3. 231-245.
- Benincà, Paola; Salvi, Gianpaolo & Frison, Lorenza 1988. L'ordine dei costituenti della frase e le costruzioni marcate. In Renzi, Lorenzo (ed.), *Grande grammatica italiana di consultazione. La frase. I sintagmi nominali e preposizionali*. Vol. 1. Bologna: Il Mulino. 115-225.
- Berretta, Monica 1994. Ordini marcati dei costituenti di frasi in italiano. La frase scissa. *Vox Romanica* 53. 79-105.
- Berretta, Monica 2002. *Quello che voglio dire è che*: le scisse da strutture topicalizzanti a connettivi casuali. In Beccaria, Gianluigi & Marengo, Carla (eds.), *La parola al testo. Scritti per Bice Mortara Garavelli*. Alessandria: Dell'Orso. 15-31.
- Berruto, Gaetano 1987. *Sociolinguistica dell'italiano contemporaneo*. Roma: Nuova Italia Scientifica.
- Bettoni, Camilla & Rubino, Antonia 1996. *Emigrazione e comportamento linguistico. Un'indagine sul trilinguismo dei siciliani e dei veneti in Australia*. Galatina: Congedo.
- Blasco Ferrer, Eduardo 1984. *Storia linguistica della Sardegna*. Tübingen: Niemeyer.
- Blasco Ferrer, Eduardo; Koch, Peter & Marzo, Daniela (eds.) 2017. *Manuale di linguistica sarda*. Berlin / Boston: Mouton de Gruyter.
- Bolognesi, Roberto 2012¹. *The Phonology of Campidanian Sardinian: A self-organizing structure*. Ghilarza: NOR.
- Bolognesi, Roberto & Heeringa, Wilbert 2005. *Sardegna tra tante lingue*. Cagliari: Condaghes.
- Bosque, Ignacio & Picallo, Carme 1996. Postnominal adjectives in Spanish DPs. *Journal of Linguistics* 32,2. 349-385.
- Calaresu, Emilia & Pisano, Simone 2017. L'italiano in Sardegna. In Blasco Ferrer *et al.* (eds.) 2017. 200-216.
- Casadei, Federica 1995. Per una definizione di "espressione idiomatica" e una tipologia dell'idiomatico in italiano. *Lingua e stile* 30,2. 335-358.
- Casadei, Federica 1996. *Metafore ed espressioni idiomatiche*. Roma: Bulzoni.
- Chafe, Wallace 1985. Linguistic differences produced by differences between

- speaking and writing. In Olson, David; Hildvard, Andrea & Torrence, Nancy (eds.) *Literacy, language, and learning*. Cambridge: Cambridge University Press. 105-123.
- Chiari, Isabella & De Mauro, Tullio 2012. The new basic vocabulary of Italian: Problems and methods. *Statistica applicata* 22,1. 21-35.
- Contini, Michele 1987. *Etude de géographie phonétique et de phonétique instrumentale du sarde*. Alessandria: Dell'Orso.
- D'Achille, Paolo; Proietti, Domenico & Viviani, Andrea 2005. La frase scissa in italiano: aspetti e problemi. In Korzen, Iørn & D'Achille, Paolo (eds.), *Tipologia linguistica e società. Due giornate italo-danesi di studi linguistici. Roma, 27-28 novembre 2003*. Firenze: Franco Cesati. 249-279.
- Dal Negro, Silvia & Fiorentini, Ilaria 2014. Reformulation in bilingual speech: Italian cioè in German and Latin. *Journal of Pragmatics* 74. 94-108.
- De Mauro, Tullio 1999a. *Grande Dizionario Italiano dell'uso (GRADIT)*. Torino: UTET.
- De Mauro, Tullio 1999b. Postfazione al Grande Dizionario Italiano dell'Uso. In De Mauro, Tullio (ed.), *Grande Dizionario Italiano dell'uso (GRADIT)*. Torino: UTET. Vol. 6. 1163-1211.
- DES = Wagner 2008.
- Dettori, Antonietta 2002. La Sardegna. In Cortelazzo, Manlio; Marcato, Carla; De Blasi, Nicola & Clivio, Gianrenzo P. (eds.), *I dialetti italiani. Storia, struttura, uso*. Torino: UTET. 897-958.
- Dettori, Antonietta 2017. Superstrato piemontese. In Blasco Ferrer *et al.* (eds.) 2017. 184-199.
- Dorian, C. Nancy 1973. Grammatical change in a dying dialect. *Language* 49,2. 413-438.
- Fillmore, Charles; Kay, Paul & O'Connor, Mary 1988. Regularity and idiomaticity in grammatical constructions: The case of *let alone*. *Language* 64. 501-538.
- Fiorentini, Ilaria 2017. Italian discourse markers and modal particles in contact. In Fedriani, Chiara & Sansò, Andrea (eds.), *Pragmatic Markers, Discourse Markers and Modal Particles. New Perspectives*. Amsterdam: John Benjamins. 419-438.
- Fischer, Kerstin (ed.) 2006. *Approaches to discourse particles*. Amsterdam: Elsevier.
- Gaidolfi, Susanna 2017a. L'italianizzazione del sardo. In Blasco Ferrer *et al.* (eds.) 2017. 476-494.
- Gaidolfi, Susanna 2017b. *Die Italianisierung des Sardischen*. Stuttgart: Steiner.
- Gargiulo, Marco 2009. Gli usi del dialetto in Sardegna. In Marcato, Gianna (ed.), *Dialetto. Usi, funzioni, forma*. Padova: Unipress. 34-43.
- Gargiulo, Marco 2011. La questione delle lingue in Sardegna. *Arena Romanistica* 8. 198-215.
- Gargiulo, Marco 2014. Lingua sarda a scuola e atteggiamento linguistico. In Marcato, Gianna (ed.), *Le mille vite del dialetto*. Padova: CLEUP. 365-372.
- Goldberg, Adele E. 1995. *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago: University of Chicago Press.
- Grassi, Corrado; Sobrero, Alberto & Telmon, Tullio 1997. *Fondamenti di dialettologia italiana*. Roma / Bari: Laterza.

- Haller, Hermann W. 1997. The dialects abroad. In Maiden, Martin & Parry, Mair (eds.), *The dialects of Italy*. London: Routledge. 401-411.
- Haller, Hermann W. 2006. *Tra Napoli e New York. Le macchiette italo-americane di Eduardo Migliaccio*. Roma: Bulzoni.
- Halliday, Michael Alexander & Hassan, Ruqaiya 1989. *Language, Context, and Text: Aspects of Language in a Social-Semiotic Perspective*. Victoria: Deakin University Press.
- Heine, Bernd & Kuteva, Tania 2005. *Language Contact and Grammatical Change*. Cambridge: Cambridge University Press.
- Hickey, Raymond 2010. Language Contact: Reconsideration and Reassessment. In Hickey, Raymond (ed.), *The Handbook of Language Contact*. Oxford: John Wiley and Sons. 1-28.
- Hilpert, Martin 2014. *Construction Grammar and its Application to English*. Edinburgh: Edinburgh University Press.
- Krefeld, Thomas 2017. Fonetica, fonologia, prosodia: diacronia. In Blasco Ferrer *et al.* (eds.) 2017. 320-338.
- Kupisch, Tanja 2021. Heritage Languages in Europe. In Montrul, Silvina & Polinsky, Maria (eds.), *The Cambridge Handbook of Heritage Languages and Linguistics*. Cambridge: Cambridge University Press. 45-68.
- Lai, Rosangela 2022a. Sardinian. In Gabriel, Christoph; Gess, Randall & Meisenburg, Trudel (eds.), *Manual of Romance Phonetic and Phonology*. Berlin / Boston: Mouton de Gruyter. 597-627.
- Lai, Rosangela 2022b. Quali politiche linguistiche per una lingua minoritaria in pericolo? In Marzo, Daniela; Pisano, Simone & Viridis, Maurizio (eds.), *Per una pianificazione del plurilinguismo in Sardegna*. Cagliari: Condaghes. 37-56.
- Lavinio, Cristina 1975. *L'insegnamento dell'italiano. Un'inchiesta campione in una scuola sarda*. Cagliari: EDES.
- Lavinio, Cristina 2019. Aspetti grammaticali dell'italiano regionale di Sardegna. *Studi di Grammatica Italiana* 36. 201-234.
- Linzmeir, Laura 2019. *Compendium of the Sassarese language: A Survey of Genesis, Structure, and Language Awareness*. Munich: Ibykos Verlag.
- Loi Corvetto, Ines 1983. *L'italiano regionale di Sardegna*. Bologna: Zanichelli.
- Loi Corvetto, Ines 1992. La Sardegna. In Bruni, Francesco (ed.), *L'italiano nelle regioni*. Torino: UTET. 875-917.
- Loi Corvetto, Ines 1993. La Sardegna. In Loi Corvetto, Ines & Nesi, Annalisa (eds.), *La Sardegna e la Corsica*. Torino: UTET. 3-205.
- Matras, Yaron 2010. Contact, Convergence, and Typology. In Hickey, Raymond (ed.), *The Handbook of Language Contact*. Oxford: John Wiley and Sons. 66-85.
- McEnery, Tony & Gabrielatos, Costas 2006. English Corpus Linguistics. In Aarts, Bas & McMahon, April (eds.), *The Handbook of English Linguistics*. Oxford: Blackwell Publishing. 33-71.
- Mithun, Marianne 2012. Exuberant Complexity: The Interplay of Morphology, Syntax, and Prosody in Central Alaskan Yup'ik. *Linguistic Discovery* 10,1. 5-26.
- Molinu, Lucia 2017. Fonetica, fonologia, prosodia: sincronia. In Blasco Ferrer *et al.* (eds.) 2017. 339-358.
- Montrul, Silvina 2016. *The Acquisition of Heritage Language*. Cambridge:

- Cambridge University Press.
- Paulis, Giulio 1983. *Lingua e cultura nella Sardegna bizantina. Testimonianze linguistiche dell'influsso greco*. Sassari: L'Asfodelo.
- Paulis, Giulio 2017. Greco e superstrati primari. In Blasco Ferrer *et al.* (eds.) 2017. 104-118.
- Pietsch, Lukas 2008. Prepositional aspect constructions in Hiberno-English. In Siemund, Peter & Kintana, Noemi (eds.), *Language Contact and Contact Languages*. Amsterdam / Philadelphia: John Benjamins. 213-236.
- Pinto, Immacolata 2012. The influence of loanwords on Sardinian word formation. In Vanhove, Martine; Stolz, Thomas; Urdze, Aina & Otsuka, Hitomi (eds.), *Morphologies in Contact*. Berlin: Akademie-Verlag. 227-245.
- Pinto, Immacolata 2015. La derivazione in sardo medievale: una prima analisi in prospettiva sociolinguistica. In Molinelli, Piera & Putzu, Ignazio Efisio (eds.), *Modelli epistemologici, metodologie della ricerca e qualità del dato. Dalla linguistica storica alla sociolinguistica storica*. Milano: Franco Angeli. 264-281.
- Piredda, Noemi 2013. *Gli italiani locali di Sardegna. Uno studio percettivo*. Frankfurt am Main: Peter Lang.
- Pisano, Simone 2015. Importanza del superstrato catalano, castigliano e italiano in alcune varietà sarde moderne: aspetti lessicali, fonetici e morfosintattici. In Filipponio, Lorenzo & Seidl, Christian (eds.), *Le lingue d'Italia e le altre. Contatti, sostrati e superstrati nella storia linguistica della Penisola*. Milano: Franco Angeli. 149-165.
- Pisano, Simone 2017. Lessico e formazione delle parole: diacronia. In Blasco Ferrer *et al.* (eds.) 2017. 397-412.
- Pisano, Simone; Piunno, Valentina & Ganfi, Vittorio 2022. Appunti per un corpus di sardo multimediale. In Marzo, Daniela; Pisano, Simone & Virdis, Maurizio (eds.), *Per una pianificazione del plurilinguismo in Sardegna*. Cagliari: Condaghes. 147-164.
- Piunno, Valentina 2018. *Sintagmi preposizionali con funzione aggettivale e avverbale*. Munich: Lincom Europa.
- Prifti, Elton 2013. *Italoamericano*. Berlin / Boston: Mouton de Gruyter.
- Putzu, Ignazio Efisio 2011. La posizione linguistica del sardo nel contesto mediterraneo. In Stroh, Cornelia (ed.), *Neues aus der Bremer Linguistikwerkstatt*. Bochum: Brockmeyer. 175-205.
- Rindler Schjerve, Rosita 1987. *Sprachkontakt auf Sardinien. Soziolinguistische Untersuchung des Sprachenwechsels in ländlichen Bereich*. Tübingen: Narr.
- Rindler Schjerve, Rosita 1998. Codeswitching as an indicator for language shift? Evidence from Sardinian-Italian bilingualism. In Jacobson, Rodolfo (ed.), *Codeswitching Worldwide: Trends in Linguistics*. Berlin / New York: Mouton de Gruyter. 221-246.
- Rindler Schjerve, Rosita 2000. Inventario analitico delle attuali trasformazioni del sardo. *Revista de Filología Románica* 17. 229-246.
- Rindler Schjerve, Rosita 2017. Sociolinguistica e vitalità del sardo. In Blasco Ferrer *et al.* (eds.) 2017. 31-44.
- Silva-Corvalán, Carmen 1994. *Language contact and change: Spanish in Los Angeles*. Oxford: Oxford University Press.
- Siemund, Peter 2008. Constraints and common paths of contact-induced lan-

- guage change. In Siemund, Peter & Kintana Noemi (eds.), *Language Contact and Contact Languages*. Amsterdam / Philadelphia: John Benjamins. 3-11.
- Simone, Raffaele 2006. Nominales sintagmáticos y no-sintagmáticos. In De Miguel, Elena; Palacios, Azucena & Serradilla, Ana (eds.), *Estructuras léxicas y estructuras del léxico*. Berlin: Peter Lang. 221-241.
- Simone, Raffaele 2007. Constructions and categories in verbal and signed languages. In Pizzuto, Elena; Pietrandrea, Paola & Simone, Raffaele (eds.), *Verbal and Signed Languages. Comparing Structures, Constructs, and Methodologies*. Berlin / New York: Mouton de Gruyter. 198-252.
- Simone, Raffaele & Masini, Francesca 2014. Light Nouns. In Simone, Raffaele & Masini, Francesca (eds.), *Word Classes*. Amsterdam / New York: John Benjamins. 51-74.
- Thomason, Sarah 2001. *Language Contact: An Introduction*. Washington: Georgetown University Press.
- Thomason, Sarah & Kaufman, Terence 1988. *Language Contact, Creolization and Genetic Linguistics*. Berkeley: University of California Press.
- Tognini-Bonelli, Elena 2001. *Corpus linguistics at work*. Amsterdam / Philadelphia: John Benjamins.
- Toso, Fiorenzo 2017. Superstrato toscano e ligure. In Blasco Ferrer *et al.* (eds.) 2017. 135-149.
- Turchetta, Barbara 2005. *Il mondo in italiano. Varietà ed usi internazionali della lingua*. Roma: Laterza.
- Virdis, Maurizio 1978. *Fonetica storica del dialetto campidanese*. Cagliari: Della Torre.
- Virdis, Maurizio 2017. Superstrato spagnolo. In Blasco Ferrer *et al.* (eds.) 2017. 168-183.
- Wagner, Max Leopold 1997². *La lingua sarda. Storia, spirito e forma*. Nuoro: Ilisso.
- Wagner, Max Leopold 2008². *Dizionario Etimologico Sardo* (a cura di Giulio Paulis). Nuoro: Ilisso.
- Weinreich, Uriel 1953. *Languages in Contact: Findings and Problems*. New York: Linguistic Circle of New York.
- Winford, Donald 2003. *An Introduction to Contact Linguistics*. Oxford: Blackwell.