

*“Le lingue sono fatte da chi le parla”*

*P. B.*

## Indice

<i>Introduzione</i> .....	3
Capitolo I .....	4
La varietà gallurese .....	4
I. 1 Cenni storici .....	5
I. 2 I Corsi in Gallura .....	10
I. 3 Il Gallurese .....	13
I. 4 La cultura gallurese .....	25
Capitolo II .....	31
La tutela delle minoranze linguistiche .....	31
II. 1 La normativa europea .....	33
II. 2 La normativa italiana .....	38
II. 3 Il ruolo della linguistica computazionale .....	46
Capitolo III .....	56
La costruzione di un dizionario-macchina gallurese con Nooj .....	56
III. 1 La linguistica e la lessicografia computazionale .....	57
III. 2 I dizionari .....	58
III. 2 Nooj .....	76
III. 3 Il dizionario gallurese .....	81
III. 4 Ulteriori considerazioni .....	84
<i>Conclusioni</i> .....	85
<i>Bibliografia</i> .....	86
<i>Sitografia</i> .....	92

## *Introduzione*

Gli studi finora trattati sulla varietà gallurese hanno esposto differenti schemi teorici sui quali si sono poi basati i trattati dei ricercatori e dei cultori dei dialetti, o presunti tali, sardo-còrsi.

Il presente elaborato vuole sfatare alcuni miti sulla varietà gallurese e sul concetto stesso di varietà o dialetto, vuole inoltre elaborare nuove strategie pratiche e applicative sull'uso dei nuovi sistemi informatici per la rivalutazione e la protezione delle lingue minoritarie.

Nel particolare il progetto che verrà esposto fornirà alcune soluzioni applicative all'unione di teorie informatiche nascenti come crowdsourcing, cloudcomputing e l'utilizzo dei dizionari macchina come mezzo di conservazione dei patrimoni linguistici.

La teoria è stata applicata nello specifico alla costruzione di un dizionario macchina gallurese con Nooj, applicazione informatica che consente la costruzione di dizionari elettronici e grammatiche da parte di tutti gli utenti, è un open source quindi disponibile anche per comunità linguistiche che non hanno possibilità di investimenti spendibili nella salvaguardia della lingua.

La prima sezione del presente elaborato tratterà della documentazione storica di quello che è il percorso evolutivo della varietà gallurese, seguirà poi una parte dedicata all'esposizione di quelle che sono le nuove scoperte informatiche in ambito di ricerca sull'unione tra linguistica computazionale e un sostegno per le lingue minoritarie, in conclusione verrà mostrato il processo di costruzione del dizionario e l'applicazione dello stesso alla lingua gallurese.

# Capitolo I

## La varietà gallurese

La varietà gallurese fa parte del panorama linguistico sardo-còrso all'interno del quale si trova anche la varietà sassarese. I dialetti sardo-còrsi sono in concorrenza con le varietà della Corsica poiché i dialetti corsòfoni in Sardegna sono parlati da circa duecentomila utenti e rappresentano la più numerosa alloglossia sarda. La maggiore città corsòfona non si trova in Corsica ma in Sardegna: il dialetto di Sassari, infatti, è parlato da 60.000-65.000 utenti, circa il doppio rispetto alla popolazione corsòfona di Ajaccio e Bastia<sup>1</sup>. Le varietà còrse sono rappresentate dai dialetti oltremontano, cismontano, ajaccino e capocorsino, situate nella parte meridionale della Corsica. Si può parlare di sostanziale unità, come nel caso del dialetto gallurese, ma esistono comunque delle sfumature che creano alcune difficoltà per quanto riguarda la scelta di un vocabolario comune e di un'ortografia unitaria<sup>2</sup>. La varietà oltremontana che maggiormente ha caratterizzato il gallurese, mostra aspetti più conservativi rispetto a quella cismontana, avendo subito più l'influsso del toscano. Il sassarese, invece, soprattutto sul piano fonologico, costituisce una varietà più vicina ai dialetti di Ajaccio e del suo entroterra.

Il presente capitolo ha l'obiettivo di ricostruire l'ambiente sociale e storico nel quale si formò in particolare il dialetto gallurese e inquadrare le sfumature dinamiche della sua evoluzione che non ha ancora trovato una definizione univoca. Il fine ultimo è quindi quello di dare una collocazione storica alle migrazioni còrse in Gallura e far luce sulla loro influenza linguistica e culturale<sup>3</sup>.

---

<sup>1</sup> M. MAXIA in TOSO F., *La Sardegna che non parla sardo*, CUEC Editrice, 2012.

<sup>2</sup> Recentemente la Consulta Intercomunale Gallurese ha preso in esame l'opportunità di avanzare una proposta ortografica per il gallurese. La proposta di unificazione dei modelli ortografici risponde all'esigenza di agevolare il lettore nella comprensione del suono e indirizzarlo in modo univoco verso un fonema identificabile da un solo segno (per ulteriori approfondimenti cfr. A cura di CONSULTA INTERCOMUNALE GALLURESE, *Gallura. Cenni storici e diversità linguistiche*, Olbia, Editrice Taphros, 2003).

<sup>3</sup> M. MAXIA, *Studi sardo-corsi*, Olbia, Editrice Taphros, 2008, p. 181 e segg.

## I. 1 Cenni storici

Nel presente paragrafo verrà fornita una traccia al fine di stabilire il contesto storico in cui si collocano gli sviluppi delle occupazioni e degli insediamenti in territorio gallurese.

Le civiltà originarie della Sardegna furono quella proto-sarda e successivamente quella nuragica, ma l'isola si colloca a livello geografico nell'area delle più importanti rotte commerciali del Mediterraneo e sin dalla preistoria fu scalo per il passaggio di diversi popoli migratori: i Balari, gli Jolesi o Iliesi, i Tibulati, forse i Libici, i Fenici, i Cartaginesi. Fu terra di conquista romana, genovese, pisana e aragonese e meta di forti migrazioni in particolare dalla vicina Corsica.

I Fenici erano presenti sulle coste sarde già nel Mille a.C. in scali stagionali: essi costruirono approdi intermedi sulle rotte dei loro commerci e fondarono diverse città. Quando nel 509 a.C. i Sardi attaccarono le città costiere, i Fenici per difendersi chiesero aiuto a Cartagine. I Cartaginesi sconfissero i Sardi e iniziarono a esercitare il loro dominio senza soggiogare la Barbagia e le zone del Logudoro della Gallura<sup>4</sup>.

Dopo la 1<sup>a</sup> Guerra Punica (238 a.C.) i Romani presero possesso della Sardegna e della Corsica togliendole a Cartagine. Gli abitanti delle città della costa, consci delle fruttuose possibilità di commercio derivanti dalla nuova occupazione, decisero di non opporsi. Le popolazioni interne, Corsi e Balari, continuarono a resistere. In Gallura i Romani trovarono uno spaccato economico e sociale multietnico: da una parte c'era Olbia, nella zona costiera, dove convivevano Punici e Sardi in condivisione di lingua, religione e attività commerciali; dall'altra la zona del Monte Limbara, dove intorno ai nuraghi veniva praticato l'allevamento. Questa divisione si accentuò durante la dominazione romana. Roma penetrò nelle zone interne anche con le strade: queste collegavano la costa alla zona montuosa del Limbara e mettevano il territorio sotto il controllo delle milizie. L'eredità romana più duratura riguarda la lingua: la latinizzazione ha lasciato un'impronta durevole mentre in Corsica e in Sardegna, il greco, l'etrusco e il fenicio-punico hanno lasciato poche tracce sulla parlata odierna. Il sostrato precedente venne coperto e trasformato dal latino, da tale commistione nacque

---

<sup>4</sup> A cura di CONSULTA INTERCOMUNALE GALLURESE, *Gallura. Cenni storici e diversità linguistiche*, Olbia, Editrice Taphros, 2003, pp. 6-7.

il sardo, parlato anche in Gallura fino al 1600 e oltre, insieme al catalano e allo spagnolo<sup>5</sup>.

Dato che i contatti tra le due sponde delle Bocche di Bonifacio erano iniziati già dal Neolitico probabilmente al momento della conquista romana i Corsi, originari della Corsica e distinti dagli Ilienses e dai Balari, si trovavano nell'estremo nord dell'isola. Si ipotizza che il loro linguaggio fosse diverso da quello degli altri popoli della Sardegna e che il gallurese si sia sviluppato successivamente come lingua romanza autoctona, assumendo le vesti di dialetto locale<sup>6</sup>.

Col declino dell'Impero le coste sarde furono protagoniste di violente incursioni barbare e saracene e questo fu il primo passo verso lo spopolamento delle zone litoranee. Furono due repubbliche marinare a liberare la Sardegna dalle invasioni e a esercitare un'influenza di tipo politico, economico e culturale. Genova si distinse nella lotta contro i Saraceni mentre Pisa fu la prima ad affermare la sua supremazia sul Cagliaritano e sulla Gallura mantenuta fino alla guerra contro gli Aragonesi. Genova intorno al Trecento estese il suo dominio su tutta l'isola e successivamente nel territorio di Sassari; i Pisani, al contrario, dovettero limitare la loro sfera di controllo alla Gallura data la presenza degli Aragonesi sino alla pace del 1326. Nel 1297, infatti, papa Bonifacio VIII aveva concesso a Giacomo II, re della Corona d'Aragona, un presunto *Regnum Sardiniae et Corsicae*: difatti gli Aragonesi incontrarono una dura opposizione e in Corsica non andarono oltre il possesso di Bonifacio. Gli studiosi concordano sul fatto che il periodo aragonese (1323-1479) fu un'epoca di profonda crisi: diminuirono i villaggi, specie nelle zone costiere, in quelle collinari e di pianura, e la popolazione continuò a spostarsi nelle regioni interne. Questo fu causato dalla lunga guerra di conquista, dal malgoverno aragonese e dalla malaria, dalle carestie e dalla pestilenza, tra cui quella memorabile del 1376 che decimò intere popolazioni<sup>7</sup>.

Fu proprio tra la fine del periodo pisano e quello aragonese che crebbe il ruolo di Tempio, fulcro della cultura gallurese. *Villa Templi* in epoca giudicale attrasse le popolazioni dei villaggi abbandonati trovandosi al centro della Gallura ed essendo facilmente difendibile. Gli ampi territori che si estendevano sino alle coste appartenevano agli abitanti della villa, proprietari o pastori, dilatando così i confini di

---

<sup>5</sup> T. PANU, *Storia di Tempio e della Gallura*, Muros (SS), Nuova Stampa Color, 2010, p. 20 e segg.

<sup>6</sup> Cfr. W. BELLODI, *Ilienses, Balari e Corsi antiche popolazioni della Sardegna*, in "Sardegna Mediterranea", n. 20 (2006): pp. 66-73.

<sup>7</sup> T. PANU, op. cit., 2010, pp. 99-100.

quello che in seguito sarebbe diventato il comune di Tempio Pausania. Questa era una delle città del regno meglio costruite e il suo territorio era molto vasto e ricco di pascoli. La pastorizia era l'occupazione più diffusa: durante l'inverno i pastori si stabilivano vicino alle coste e si spostavano verso Tempio con le loro famiglie durante i mesi più caldi. Successivamente i terreni disponibili, prima abbandonati per il nomadismo dei pastori, con il progressivo sviluppo demografico furono delimitati e venne stabilita la proprietà della terra. Nel mondo della pastorizia esistevano status sociali diversi: i signori del bestiame, ossia grandi proprietari terrieri che avevano alle proprie dipendenze i servi pastori; vi erano poi i pastori liberi, piccoli proprietari di un gregge o di una mandria. Questi ultimi favorirono il riconoscimento di un ulteriore status sociale: i piccoli proprietari che gestivano direttamente l'impresa rurale o pastorale e cominciarono a produrre non soltanto per la sussistenza, ma anche per il mercato<sup>8</sup>.

I Corsi nel frattempo si battevano contro i Genovesi, i quali furono costretti dalla resistenza a cedere l'isola alla Francia. La Corsica si unì a quest'ultima il 30 novembre 1789. La Sardegna, invece, col trattato di Londra (1718) e dell'Aia (1720) fu assegnata ai re sabaudi e da allora assunsero l'appellativo di re di Sardegna. Col governo sabauda iniziò il ripopolamento programmato che però non era in grado di soddisfare le esigenze socio-culturali dell'isola, di conseguenza il progresso fu lento e contenuto.

La Corsica divenuta francese andò incontro a un periodo di grave crisi: la mancanza di risorse e la fame spinsero i Corsi a fare della Sardegna una fonte di approvvigionamento. Fiorì allora il contrabbando: il traffico di grano, di bestiame e di prodotti caseari si svolgeva tra le due rive. L'arcipelago maddalenino divenne terra di pascolo in attesa del trasferimento del bestiame sul litorale di Bonifacio. Questa, in un primo tempo, fu la causa dell'insediamento dei pastori còrsi a La Maddalena<sup>9</sup>.

La soppressione del contrabbando fu il motivo, o pretesto ufficiale, per occupare militarmente l'Arcipelago di La Maddalena. Così il 14 ottobre 1767 ebbe inizio, da parte dell'esercito sardo-piemontese, l'insediamento nelle isole, abitate da poche famiglie di pastori còrsi<sup>10</sup>. Questi dovettero riconoscere la sovranità politica del Regno

---

<sup>8</sup> T. PANU, op. cit., 2010, pp. 102-125-126-127.

<sup>9</sup> R. DE MARTINO, *Il Dialetto Gallurese*, in a cura di SALVATORE BRANDANU, *La Gallura, una Regione diversa in Sardegna*, San Teodoro, Editrice I.C.I. Mar, 2001, pp. 162-165-166.

<sup>10</sup> Cfr. S. RATTU, *S. Teresa di Gallura*, in *Studi storici in onore di F. Loddo Canepa*, Bologna, 1956, vol. I, pp. 290 e segg.

di Sardegna e integrarsi con gli immigrati galluresi, pur conservando le proprie usanze e la parlata di Bonifacio<sup>11</sup>.

Il ripopolamento della Gallura marittima avvenne invece per opera dei Corsi immigrati e delle famiglie provenienti dai paesi dell'interno, soprattutto da Tempio e da Aggius. Lo stesso fenomeno si ebbe intorno alle chiese campestri dove sorsero i primi nuclei di centri abitati: tra i villaggi i Corsi si stabilirono piccoli nuclei familiari. Fu una colonizzazione rada e dispersa nel territorio, ma numerosa e costante. L'emigrazione corsa fu favorita dalla vicinanza delle due isole e dall'esigua difficoltà di inserimento in terre pressoché disabitate. Questa sorta di usucapione trovò legittimazione nell'Editto delle Chiudende (1820), esteso a tutta l'isola, il quale però favorì solamente i ricchi possidenti, gli unici in grado di recintare in fretta ampie aree del territorio. Per questa ragione l'Editto fu sospeso nel 1833 da Carlo Alberto. Il noto Editto in Gallura ebbe scarsissima applicazione in quanto era già stato messo in atto: prima del Settecento infatti, i pastori erano diventati proprietari di grandi terreni sottratti al feudatario, al signore, alla villa, o al demanio regio. Tale appropriazione portò alla nascita dei pastori-proprietari, passando da un regime di proprietà comune delle terre a un regime di proprietà privata non senza scontri.

La pirateria divenuta una piaga fu affrontata risolutamente a livello internazionale. Liberati dal pericolo algerino da parte di Inglesi e Olandesi prima (1816), e Francesi poi (1830), i Sardi iniziarono a ripopolare le zone marittime, incrementando lentamente l'industria, l'agricoltura e il commercio. Le attività marinare restarono a lungo nelle mani dei continentali, specie dei napoletani, che praticavano soprattutto la pesca del corallo, del pesce azzurro e delle aragoste. Sino al XII secolo erano stati i Liguri a occuparsene, ma in seguito alla dominazione spagnola sulla Sardegna e sull'Italia meridionale s'intensificò la presenza dei pescatori campani: molti di loro si stabilirono definitivamente a Golfo Aranci, a Santa Teresa e alla Maddalena, contribuendo al ripopolamento della zona nord-orientale dell'isola<sup>12</sup>.

Nel corso dell'Ottocento continuarono a svilupparsi le colonizzazioni guidate e l'insediamento rurale mentre la popolazione dell'interno iniziava a occupare le zone costiere galluresi. Verso gli inizi del Novecento avvenne il passaggio dall'insediamento sparso a quello accentrato attraverso la graduale concentrazione degli abitanti degli stazzi. Da questi due fenomeni è nata la Gallura contemporanea, caratterizzata dal

---

<sup>11</sup> T. PANU, op. cit., 2010, p. 201.

<sup>12</sup> R. DE MARTINO, op. cit., 2001, pp. 166-167.



preponderante sviluppo dei centri costieri. I comuni dei territori collinari restavano più numerosi rispetto alla zona litoranea ma la discesa verso le zone costiere era in aumento. Per quanto riguarda lo sviluppo demografico, fin dalla metà dell'Ottocento i comuni di montagna e di collina interna manifestavano un trend di crescita contenuto rispetto alle altre zone mentre i comuni di collina litoranea e di pianura mostravano una crescita più accelerata.

Dopo la Seconda guerra mondiale in Gallura ci furono tre importanti cambiamenti: l'autonomia amministrativa delle antiche frazioni, la quale ridusse i confini del territorio comunale; lo sviluppo turistico delle zone costiere, in particolare Olbia, che tolse alla città di Tempio il ruolo di guida economica della Gallura; la crisi dell'economia degli stazzi, che indusse la popolazione delle campagne a stabilirsi in città.

La classe dirigente di Tempio del Secondo dopoguerra, consapevole del cambiamento economico e sociale, pensò di affrontare quest'ultimo potenziando il ruolo della città come centro di servizi: qui avevano sede gli uffici pubblici di competenza territoriale, le scuole superiori, esercizi commerciali e la cattedra episcopale. L'agricoltura aveva un ruolo socio-economico marginale: era praticato l'allevamento semi-brado di bovini per la produzione di carne. Il modello antico dello stazzo con la famiglia coltivatrice diretta iniziava a essere abbandonato.

Il periodo compreso tra il 1962 e il 1984 è considerato il Rinascimento gallurese. Nel 1962 furono costituiti dalla giunta regionale i comitati zonal di sviluppo per il Piano di Rinascita e la Gallura fu divisa in due zone: Olbia e Tempio; fu come erigere una barriera fra le «due Gallure». I comitati zonal svolsero un ruolo consultivo ma ebbero una funzione importante di coordinamento fra le diverse istanze locali<sup>13</sup>.

Lo sviluppo fu organizzato per poli: la scelta fu quella di un percorso industriale. Il polo più debole era quello di Olbia, anche se il turismo stava iniziando a trasformare l'economia e la fisionomia territoriale della Gallura costiera<sup>14</sup>. Il 28 settembre 1961 a Olbia, il principe Karim Aga Khan illustrò ai giornalisti il progetto della Costa Smeralda, e il 10 agosto 1964 fu inaugurato il porto turistico di Porto Cervo. Il turismo ha trasformato e modernizzato l'economia delle coste attraverso il potenziamento dell'edilizia senza però estendersi alle restanti attività produttive.

---

<sup>13</sup> T. PANU, op. cit., 2010, p.369.

<sup>14</sup> Ivi, p. 380.

La valorizzazione delle coste e dell'interno era fortemente squilibrata. La rinascita economica si concentrava solo nei poli industriali costieri a discapito delle zone interne. Lo sviluppo della Gallura ebbe i suoi aspetti positivi, ma tese a un netto distacco socio-economico tra le zone dove si collocavano le industrie di base e le zone dove la pastorizia e l'agricoltura erano le sole attività economiche.

Negli anni ottanta del Novecento le attenzioni dei politici galluresi si concentrarono sul Piano di Rinascita<sup>15</sup>. La questione sull'utilizzo o la salvaguardia del territorio e il turismo come fonte principale o secondaria dello sviluppo avevano da sempre preoccupato l'ambiente politico. In Gallura la scelta si è spostata verso un uso talvolta speculativo del territorio e si è manifestata la pericolosa tendenza a fare del turismo la fonte più importante, se non unica, di reddito. Questo ha causato gravi danni ambientali e una pesante dipendenza da capitali esterni.

## I. 2 I Corsi in Gallura

I Corsi sono un popolo migratore per eccellenza<sup>16</sup>. Le migrazioni còrse verso la Sardegna si succedettero ininterrottamente dal basso medioevo e continuarono in età moderna<sup>17</sup>. Una delle ragioni principali della migrazione in Gallura è il fatto che quest'ultima, rimasta pressoché disabitata nella seconda metà del Trecento, procurava magre rendite ai feudatari catalani. In questo modo l'immigrazione dalla Corsica veniva favorita dalle autorità catalano-aragonesi nella prospettiva di incrementare le entrate attraverso l'insediamento di nuovi vassalli.

La prima documentazione sicura della presenza di pastori còrsi è documentata dal passo di un documento scritto in catalano in pieno Trecento (1358 c.ca). Si tratta del *Compartiment de Sardenya*, il quale attesta che nel salto di Cassari «los Corsos e altres homens [...] tenene aqui bestià»<sup>18</sup>. Risulta così palese che i pastori còrsi con il loro caratteristico modello insediativo, lo stazzo, non giunsero in Gallura nel Cinquecento ma nei secoli precedenti. Le fonti, infatti, prendono atto di un fenomeno preesistente al momento della documentazione<sup>19</sup>.

---

<sup>15</sup> T. PANU, op. cit., 2010, p. 381.

<sup>16</sup> Cfr. F. BRAUDEL, *Civiltà e imperi del Mediterraneo nell'età di Filippo II*, Torino, Piccola Biblioteca Einaudi, 2010.

<sup>17</sup> M. MAXIA, *I Corsi in Sardegna*, Cagliari, 2006, Edizioni Della Torre, p.10.

<sup>18</sup> P. DE BOFARULL I MASCARÓ, *Colección de documentos inéditos del Archivo de la Corona de Aragón*, XI, S. I., Establecimiento Litográfico y Tipográfico de José Eusebio Monfort, 1847-1851, p. 822.

<sup>19</sup> M. MAXIA, op. cit., 2006, p.148.

A partire dal Quattrocento le migrazioni còrse crebbero in maniera esponenziale, tanto da rendere la componente còrsa predominante lungo la fascia litoranea e nel retroterra che connette l'isola dell'Asinara al settore più settentrionale delle Baronie<sup>20</sup>. La varietà oltremontana del dialetto còrso penetrò nelle campagne e nei paesi circostanti, spingendosi anche sulla costa occidentale (Trinità, Badesi, Viddalba, Santa Maria) sino a Castelsardo (gallurese-sassarese) e a est nelle vicinanze di Olbia e ai confini della Baronia (San Teodoro e Ittireddu).

Come attesta M. Le Lannou:

esso è riuscito a penetrare nella maggior parte dei villaggi dei dintorni di Tempio mescolato a forme dell'antica lingua logudorese, ma non ha toccato affatto uno di essi, che pure è il più settentrionale, cioè Luras, i cui abitati, circondati da dialetti galluresi, sono restati completamente fedeli alla loro parlata sarda e capiscono con difficoltà i contadini degli stazzi.<sup>21</sup>

È importante sottolineare che i Corsi non imposero il loro dialetto ma che la parlata locale assunse alcuni tratti dell'oltremontano conservando molte delle proprie caratteristiche<sup>22</sup>. Il gallurese presenta infatti elementi originali anche a confronto coi dialetti còrsi più simili (quelli dell'Alta Rocca e dell'Alto Tàravo)<sup>23</sup>.

La documentazione linguistica delle varietà sardo-còrse è esigua. Questo si spiega nella percezione di tali varietà che venivano considerate, al pari del còrso dell'isola madre, come dialetti di scarso prestigio. Nel presente la loro rivalutazione sta avvenendo in controtendenza rispetto ad altre varietà italiane e allo stesso sardo che appare avviato verso un progressivo disuso (fatta eccezione per alcune zone più conservative). Sono ragioni di ordine socio-linguistico e politiche a favorire la riemersione sia del còrso che dell'alloglossia corsòfona della Gallura<sup>24</sup>.

Nella seconda metà del Cinquecento, secondo lo storico Giovanni Francesco Fara, la maggior parte della Gallura, pur essendo quasi priva di villaggi, era popolata da migliaia di pastori che dimoravano nelle campagne con le loro famiglie. Degli antichi abitati restavano solo quelli di Terranova, Aggius, Bortigiadas, Calangianus, Luras,

---

<sup>20</sup> M. MAXIA, op. cit., 2006, p. 10.

<sup>21</sup> M. LE LANNOU, *Pâtres et paysans de la Sardaigne*, "Revue de géographie alpine", Anno 1942, Volume 30, Numero 3, p. 48.

<sup>22</sup> R. DE MARTINO, op. cit., 2001, p.168.

<sup>23</sup> F. TOSO, *La Sardegna che non parla sardo*, Cagliari 2012, CUEC Editrice, p. 28.

<sup>24</sup> M. MAXIA, op. cit., 2006, pp. 19-20.

Nuchis e Tempio, ma nel territorio era presente un diffuso insediamento sparso. Gli studiosi del popolamento della Gallura non prestarono attenzione alla testimonianza del Fara e forse fraintesero le sue parole. Lo studioso, scrivendo che la Gallura era disabitata, intendeva riferirsi alla rarefazione dei centri abitati. Infatti poi precisava che i pastori conducevano migliaia di greggi: «multique ex illis pastorician et agresten cum tota familia in montibus degunt vitam, mille greges illi totidemque armenta per herbas pascunt»<sup>25</sup>.

Si può quindi escludere fermamente che la migrazione còrsa sia iniziata soltanto alla fine del Cinquecento o durante il Seicento, come credeva Le Lannou<sup>26</sup>. Il popolamento delle campagne costituiva una realtà ben radicata<sup>27</sup>, e si può dare per certo che nel Cinquecento la Gallura, compresi i villaggi veri e propri, fosse abitata soprattutto da genti di origine còrsa. Questo si può dedurre dal forte influsso esercitato dal sassarese e dal gallurese nei confronti del logudorese. Il fatto più eclatante è rappresentato dalla formazione del dialetto logudorese settentrionale; le opinioni degli studiosi appaiono ancora una volta controverse: essi attribuirono all'influsso toscano e continentale l'origine dei fenomeni fonetici e lessicali che lo caratterizzano. Non è stato preso in considerazione un dato importante: l'influsso italiano nella Sardegna settentrionale cessò, per quanto riguarda il toscano, entro la prima metà del Trecento, e quello genovese entro la metà del Quattrocento. È difficilmente ipotizzabile che le correnti culturali italiane che fino al Cinquecento toccarono esigue fasce sociali potessero influenzare la lingua del popolo<sup>28</sup>.

Lo spopolamento di gran parte della Gallura e conseguentemente la produzione di un irrisorio numero di documenti impediscono purtroppo di affrontare la questione in modo risolutivo, tuttavia gli elementi di cui si dispone appaiono sufficienti per retrodatare la vigenza di questa varietà almeno alla metà del Quattrocento<sup>29</sup>. Ad esempio le forme documentate nei registri parrocchiali di Tempio riflettono una pronuncia còrso-gallurese; i cognomi attestati in questa fonte non hanno subito l'italianizzazione dell'Ottocento. Sul piano linguistico questo dato testimonia che la maggioranza degli abitanti di Tempio e del suo vasto territorio già agli inizi del

---

<sup>25</sup> F. FARA, *Chorographia Sardiniae (1580-85)*, Opera, Vol I, a cura di E. CADONI, Sassari, S.n., 1992.

<sup>26</sup> M. LE LANNOU, op. cit., 1942, p.48.

<sup>27</sup> M. MAXIA, op. cit., 2008, p.81.

<sup>28</sup> M. MAXIA, *Studi storici sui dialetti della Sardegna settentrionale*, Sassari, Stadium Adf, 1999, pp. 112-113.

<sup>29</sup> M. MAXIA, op. cit., 2006, p. 21.

Seicento doveva parlare una varietà di origine corsa: il sardo conservato solo dei documenti ufficiali si avviava quindi verso un uso residuale<sup>30</sup>.

Come accennato la varietà corsòfona dell'isola vicina interagì anche col sardo logudorese. Si innescarono numerosi scambi lessicali e l'influenza di una serie di fonemi che giustificano le divergenze del sassarese e del gallurese rispetto al corso dell'isola madre e la frammentazione del logudorese in alcune sottovarietà tra le quali quella settentrionale: questa si caratterizza per la presenza di numerosi corsismi lessicali talvolta interpretati come toscanismi. A seguito della presenza corsa potrebbe anche risalire l'abitudine da parte dei sassaresi e dei galluresi di chiamare *li Cossi* (i Corsi) i nuovi immigrati. Questi ultimi ormai dovevano essere percepiti dagli stessi Corsi di più antico insediamento come popolazione altra<sup>31</sup>.

Il distacco del sassarese e del gallurese dal corso rappresenta un fatto relativamente antico. Le differenti risoluzioni fonetiche come la generalizzazione dell'epitesi, la presenza di morfemi sconosciuti in Corsica, le innovazioni nella costruzione del periodo, la presenza importante di prestiti logudoresi e catalani, l'influsso prodotto sul logudorese e sul catalano di Alghero, rappresentano le prove del cambiamento che ha attraversato queste nuove parlate in un periodo che inizia nel primo Trecento e si protrae fino alla prima metà del Seicento. Questa fase ha portato alle tre varietà che fusero in modi differenti il corso e il sardo: due di tipo toscano (sassarese e gallurese) e una di tipo sardo (logudorese settentrionale). Tali dialetti dovevano già avere le sembianze odierne come testimoniato dalla maturità della lingua dei brani poetici logudoresi della fine del Cinquecento e quelli galluresi del Settecento<sup>32</sup>.

### I. 3 Il Gallurese

Il dialetto gallurese è parlato in tutta la Gallura anche se al suo interno esistono delle differenziazioni zonal di pronuncia e fenomeni lessicali. Il confine meridionale della sua area di influenza è dato dal corso inferiore del fiume Coghinas, dalla catena del Limbara e dalla linea che unisce la sua estremità orientale con Monte Nieddu (S. Teodoro) fino a comprendere le borgate galluresi di Budoni. La Gallura comprende i comuni di: Tempio, Olbia, Luras, Calangianus, Aggius, Santa Teresa, Arzachena,

---

<sup>30</sup> M. MAXIA, op. cit., 2006, pp. 150-151.

<sup>31</sup> Ivi, p. 239.

<sup>32</sup> M. MAXIA, op. cit., 1999, pp. 117-118.

Luogosanto, Palau, L'Aglièntu, La Maddalena, Trinità d'Agultu, Telti, Golfo Aranci, Santa Maria Coghinas, Badesi, Viddalba, S. Antonio di Gallura, Loiri Porto San Paolo, San Teodoro, Budoni, Erula e Padru. E altri villaggi: Bassacutena (nel comune di Tempio), Nuchis (nel comune di Tempio), San Pantaleo e Berchideddu (nel comune di Olbia), San Pasquale (nei comuni di Tempio e S. Teresa)<sup>33</sup>.

Le varie dominazioni e l'evoluzione storica della Gallura hanno fatto sì che il gallurese subisse influssi non solo còrsi, ma anche pisani, genovesi, catalani, spagnoli e piemontesi, i quali hanno avuto delle ripercussioni notevoli anche di tipo sociale, culturale e soprattutto linguistico<sup>34</sup>. È importante perciò definire quali siano i caratteri del gallurese e quali le affinità o differenze specifiche rispetto al còrso. Al momento sembra imporsi la tesi per cui il gallurese, seppur con elementi sardi, evidenzia la prevalenza di influsso còrso (oltremontano), il suo consonantismo invece è di tipo italiano<sup>35</sup>.

Il filologo I. Petnakov afferma che «la Corsica insieme alla Gallura ed a Sassari, ai dialetti meridionali del continente, alla Sardegna ed alla Sicilia riflette le condizioni linguistiche del corso oltremontano» e il sassarese «risente più forte l'influsso toscano e in modo particolare pisano, dati gli intimi rapporti che intercorrono tra Pisa e la Sardegna nei secoli XII e XIII»<sup>36</sup>. Secondo il glottologo tedesco M. L. Wagner il gallurese e il sassarese si distinguono dal sardo per morfologia e sintassi sostanzialmente italiana, mentre il lessico, di base toscana, presenta molti elementi sardi, spagnoli e genovesi (questi ultimi soprattutto nel sassarese). In accordo con Petnakov egli afferma che il gallurese «è un dialetto essenzialmente corso e si avvicina più particolarmente al dialetto oltremontano parlato nella parte meridionale della Corsica»<sup>37</sup>.

Per costruire un fondamento di tali teorie sarebbe opportuno analizzare attentamente una documentazione coerente ed elaborata sul piano diacronico e diffusa sul piano sincronico: per il sardo si dispone dei testi medioevali dei secc. XI-XIII; per il sassarese e il gallurese ci si deve attenere alle raccolte di poesie popolari e alla lingua

---

<sup>33</sup> M. MAXIA, op. cit., 2008, pp. 35-37.

<sup>34</sup> A cura di CONSULTA INTERCOMUNALE GALLURESE, op. cit., p.28.

<sup>35</sup> R. DE MARTINO, op. cit., 2001, pp. 157-158.

<sup>36</sup> I. PETNAKOV, *Appunti sui dialetti sardo-settentrionali*, in "Archivum Romanicum", XXV, 1921.

<sup>37</sup> M. L. WAGNER, a cura di G. PAULIS, *La lingua sarda*, Nuoro, Ilisso Edizioni, 1997, p. 345.

parlata. La mancanza di adeguate fonti scritte è quindi, anche se solo in parte, compensata dalla tradizione<sup>38</sup>.

L'insufficienza di fonti rappresenta un ostacolo considerevole per chi fa ricerca in questo settore. Fatta eccezione per l'epigrafe quattrocentesca di S. Vittoria del Sassu<sup>39</sup>, non si dispone di alcun testo scritto in gallurese o in sassarese se non a partire dal Seicento e nella maggior parte dei casi presentano interferenze di tipo corso in documenti scritti in sardo o in altre lingue. Si deve considerare in alcuni dei documenti redatti in una delle due varietà, il registro richiesto dalla comunicazione scritta era stereotipato, dato il condizionamento della destinazione d'uso (religioso e notarile) poco affine al parlato<sup>40</sup>.

Per quanto riguarda il gallurese la prima documentazione di tipo letterario è costituita dal corpus di poesie del religioso tempiese Gavino Pes (1724-1795), conosciuto come Don Baìgnu, la cui attività si svolse nel Settecento. La fonetica, la morfologia, la sintassi e il lessico delle composizioni dell'autore non differiscono dall'odierno gallurese comune contemporaneo che ha come centro di riferimento Tempio Pausania<sup>41</sup>. In generale tutta la letteratura gallurese fa riferimento alla varietà comune: la parlata comune è diffusa in tutta la Gallura centro-orientale e precisamente nei comuni di S. Antonio di Gallura, Luogosanto, Aglientu, S. Teresa Gallura, Palau, Telti, Loiri-Porto S. Paolo e S. Teodoro; nei comuni di Olbia e di Golfo Aranci il gallurese è diffuso maggiormente nelle campagne, mentre nei centri abitati è diffuso prevalentemente il logudorese. Nel dominio del gallurese comune rientrano anche il comune anglonese di Erula, le borgate di Campudùlimu e Modditonalza (frazioni di Perfugas) e il nord dei territori comunali di Tula, Oschiri, Berchidda, Monti e Padru; a sud di Olbia la linea di contatto tra il gallurese e il logudorese passa tra le borgate di Su Canale (gallurese), Enas (gallurese), Berchideddu (gallurese), Sa Castanza (logudorese), Mumusi (logudorese), Pedru Gaias (logudorese), Li Coddì (gallurese), Monti Littu (gallurese), Azzanì (gallurese), Vaccileddi d'Ovilò (gallurese). Il dominio gallurese arriva sino al comune di Budoni e di Torpé dove si spinge sino a Brunella e San Gavino. Il confine meridionale del gallurese è rappresentato dal torrente di denominazione gallurese *Riu di li Cappulati* (lettm. 'rio delle tartarughe')<sup>42</sup>.

---

<sup>38</sup> R. DE MARTINO, op. cit., 2001, pp. 157-158.

<sup>39</sup> Per ulteriori approfondimenti cfr. M. MAXIA, op. cit., 2008, pp. 210 e segg.

<sup>40</sup> Ibid.

<sup>41</sup> M. MAXIA, op. cit., 1999, pp. 107-108.

<sup>42</sup> M. MAXIA, op. cit., 2008, p.35.

### I.3.1 Caratteristiche specifiche della lingua gallurese

Come detto in precedenza la varietà gallurese presenta al suo interno delle seppur minime differenziazioni: in Gallura i centri abitati e il contado hanno due registri, uno alto rappresentato dalle varietà di Tempio e dei vicini centri di Calangianus, Nuchis e Aggius, l'altro basso, detto *faeddu di lu pasturíu* ('linguaggio dei pastori') cioè la varietà della popolazione che abita negli stazzi (il lessico in questo caso risulta più rustico)<sup>43</sup>.

Al riguardo lo storiografo V. Angius era del parere che

in questa provincia [la Gallura] sono uditi diversi dialetti. I terranovesi, lurisinchini, e bortigiadesi parlano il sardo; i tempiesi, agiesi, calangianesi e nuchesini il gallurese; i longonesi una lingua mista; gli isolani [maddalenini] il corso mescolato di gallurese e genovese: [...]. Le differenze che la pronuncia porta tra gli stessi galluresi sono le seguenti. L'agiese parla spedito e vibrato, il nuchesino lento, il calangianese più allungato, il tempiese tiennesi nel mezzo, e però la sua pronuncia è assai gentile<sup>44</sup>.

A seguire i caratteri distintivi delle aree galluresi<sup>45</sup>:

- *Santa Teresa Gallura*: vi si parla un gallurese tramezzato da parole còrse. Non sono presenti differenze fonetiche ma solo lessicali;
- *Calangianus*: la pronuncia è molto accentuata quindi facilmente riconoscibile. Caratteristico di questo dialetto il raddoppiamento della *m* (*fummu* per *fumu*, *semmu* per *semu*); l'elevazione della labiale dal grado medio al grado tenue (*bapu* per *babbu*); rispetto alle altre parlate ha conservato un elevato numero di arcaismi;
- *Aggius*: distintiva l'assibilazione della *c* palatale (*zincu* per *cincu*, *zentu* per *centu*, *zena* per *cena*) e la trasformazione di *c* della *chj* (*ciai* per *chjiai* = 'chiavi'); degradamento delle esplosive iniziali nelle corrispondenti sonore (*lu góri* per *lu córi*); velocità della parlata (questo fatto non è un fenomeno di natura

<sup>43</sup> M. MAXIA, op. cit., 1999, p.109.

<sup>44</sup> V. ANGIUS, in G. CASALIS, *Dizionario geografico storico statistico degli Stati di S. M. il Re di Sardegna*, vol. VII, voce Gallura, Torino, 1840, p. 141.

<sup>45</sup> Per ulteriori approfondimenti cfr. A cura di CONSULTA INTERCOMUNALE GALLURESE, op. cit., 2003, p. 33 e segg.



fonetica ma probabilmente è dato dallo spirito vivace dei parlanti). La varietà agnese si riscontra anche tra i pastori di Viddalba, Badesi e Trinità (sue antiche frazioni) e di Vignola, *cussorgia* al confine tra Aglièntu e Trinità;

- *La Maddalena*: la massiccia corrente immigratoria influì sulla stratificazione sociale e conseguentemente sul dialetto il quale principalmente còrso-genovese, risentì degli influssi del gallurese e di altre parlate continentali. Il maddalenino oggi manifesta caratteristiche eterogenee e sta subendo un forte processo di italianizzazione;
- *Golfo Aranci*: vi risiedono numerose famiglie miste che parlano, come se fosse una lingua franca, l'italiano, riducendo così le aree di diffusione delle varie parlate locali. Si rileva un'ampia presenza di ponzesi e partenopei, i quali hanno introdotto molte loro espressioni tipiche nell'uso dell'italiano che contiene tracce di maddalenino e di catalano;
- *Arzachena*: conta circa 1500 abitanti, la metà di questi è di origini e di parlata gallurese. L'uso dell'idioma gallurese è forte nelle contrade più distanti dalla costa: Abbiadori, Santa Teresina, Liscia di 'acca, Cumitoni, L'Agghjesi, Pulicinu. Nelle zone costiere il gallurese inizia a svanire a causa della tendenza italianizzante del lessico e delle strutture sintattiche. In alcuni ambiti ristretti della cittadina arzachenese il gallurese risulta ancora vitale poiché associato a uno specifico ambiente culturale.

Ai confini della Gallura si trovano alcune isole linguistiche in passato logudoresi ma attualmente bidialettali<sup>46</sup>: Olbia nel cui centro la parlata gallurese è talmente penetrata da contrastare la prevalenza del logudorese; Bortigiadas ormai divenuta quasi esclusivamente gallurese; Luras con influssi palesi di strutture galluresi nella parlata originaria<sup>47</sup>:

---

<sup>46</sup> F. CORDA, *Saggio di grammatica gallurese*, Cagliari, Editrice 3T, 1983, p.63.

<sup>47</sup> A cura di CONSULTA INTERCOMUNALE GALLURESE, op. cit., 2003, p. 32.

- *Olbia*: sino alla fine dell'Ottocento la popolazione parlava in prevalenza in logudorese ma nelle campagne la popolazione era di parlata esclusivamente gallurese. Le località vicine all'abitato erano denominate con toponimi galluresi dai pastori e logudoresi dai cittadini (*Sa Costanza* e *Rudalza* in città, *La Castagna* e *Rutagghja* in campagna). Nell'ultimo ventennio, per lo sviluppo della Costa Smeralda e per le immigrazioni dalle campagne e dagli altri centri della Gallura, del Goceano e della Barbagia, il gallurese si è diffuso capillarmente, mentre le famiglie che si sono mantenute fedeli al logudorese sono generalmente bidialettali;
- *Bortigiadas*: fino ai primi del Novecento fu un'isola linguistica logudorese in area gallurese. Attualmente vi si parla il dialetto gallurese di tipo aggeese;
- *Luras*: il logudorese è diffuso in modo maggioritario, ma i Luresi conoscono anche il dialetto gallurese che padroneggiano senza difficoltà.

Nell'illustrazione delle caratteristiche peculiari del gallurese è fondamentale mettere in luce i tratti dell'influenza còrsa a livello linguistico. Ancora una volta si dimostra pertinente citare l'Angius, questi asserisce che «la lingua gallurese non ha altra origine della corsa, e a parte alcune piccole particolarità, si assomiglia in tutto a quella de' sartenesi, tra' quali un gallurese non pare straniero»<sup>48</sup>.

L'influenza còrsa nella varietà presa in esame è dimostrata da vari fenomeni tra cui il vocalismo: nel gallurese le *e* dinanzi ad *r* (o a *l* derivato da *r*) passa generalmente, per apofonia o digradamento vocalico, ad *a*: *abbaltu*, 'aperto', *addaretu*, 'dietro' (lat. *de retro*), *arimani*, 'ieri' (lat. *heri mane*). Questo fenomeno è tipico del còrso oltremontano mentre nel cismontano si verifica il passaggio contrario *a* > *e*. Esistono limitate oscillazioni d'uso per qualche infiltrazione o scambio tra le due varietà, o semplicemente per conformazione analogica: *barba* e *berba*, *baretta* e *beretta*, *carbone* e *cherbone*. Ancora nel cismontano *a* si altera in *e* anche dinanzi ad *n*: *galenti* (oltr. *galanti*), *grendi* (grandi), *pientu* ('pianto') (esteso anche all'oltremontano)<sup>49</sup>.

Sono di influenza còrsa i casi di metafonìa e metaplasmo. La metafonìa, propria del sardo, secondo il filologo E. B. Ferrer si sarebbe diffusa per influenza del

<sup>48</sup> V. ANGIUS, op. cit., 1840.

<sup>49</sup> R. DE MARTINO, op. cit., 2001, p. 182 e segg.

logudorese e del campidanese<sup>50</sup>. Per F. Corda, invece, esso è estraneo al gallurese, ma ammette che la sua applicazione nel logudorese (nuorese) e campidanese comporta per questi dialetti le pronunce di *nòa / nou*, *mèle / Mélis*, *bèssu / béssis* e si riflette sulla pronuncia dell'italiano regionale sardo. In tal modo anche un gallurese, come tutti i Sardi, nell'uso della lingua italiana pronuncerà *Péru*, *buoni*, *ésci* (con la vocale tonica stretta essendo seguita da *i-u*) di contro a *pèra*, *buòne*, *èSCO* (con la vocale tonica larga essendo seguita da *a-e-o*)<sup>51</sup>.

Per quanto riguarda il consonantismo si considerino i seguenti fenomeni<sup>52</sup>:

- Lenizione: è legata alla tendenza ad attenuare l'intensità delle consonanti a maggiore infissione (le occlusive sorde *p*, *c*, *t* e la fricativa *f*). In gallurese porta a indebolire e a lenirle riducendole alle corrispondenti sonore (*b*, *g*, *d* e *v*) e spesso al dileguo *abbrì*, 'aprire', *piscadori*, 'pescatore', *siguru*, 'sicuro'; con dileguo *arriatu*, 'arrivato', *ciòanu*, 'giovane', *ciudda*, 'cipolla'. Il fenomeno scompare nel passaggio dal dialetto all'italiano che viene pronunciato in modo relativamente corretto, ad eccezione della chiusura delle vocali *e - o* e per una certa marcatura consonantica (non forte come quella degli altri sardi). F. Corda osserva che

nelle combinazioni dei sintagmi che non sono soggetti a raddoppiamento sintattico la *f* subisce una lenizione e digrada a *v*: *fami / la - vami*, *fenu / lu - venu*, *fiori / li - viori*. In identica posizione la *v* scompare per lenizione totale (con la comparsa dell'afèresi): *vacca / la 'acca*, *ventu / lu - 'entu*, *vinutu / éra 'inutu*. Altrettanto avviene, in certi casi, con *g(h)*: *gula / la 'ula*<sup>53</sup>.

Altra forma di lenizione è data dal passaggio *ci > sci(j)* (dall'affricata prepalatale *c* di 'certo' alla fricativa prepalatale sonora *j* del fr. *jour*), analogo a *ci > sci* del toscano (*camiscia*, *bascio*), generalmente in corrispondenza della prima trasformazione, nel neolatino, dei termini uscenti in *-sius*, *-sia*, *-sium* (*casius*, *camisia*, *basium*) es. *basgiu*, *camisgia*, *sdrisgiutu* ('sdrucito', 'scucito');

<sup>50</sup> E. BLASCO FERRER, *Storia linguistica della Sardegna*, Tubingen, Max Niemeyer, 1984, pp. 259-60.

<sup>51</sup> F. CORDA, op. cit., 1983, p.63.

<sup>52</sup> R. DE MARTINO, op. cit., 2001, p. 183 e segg..

<sup>53</sup> F. CORDA, op. cit., 1983, p. 69.

- Cacuminazione: sono dette cacuminali le doppie *-ddh* *-gghj* e *-cchj*. Es. *baeddhu*, ‘mento’, *muntinagghju*, ‘mondezzaio’, *occhj*, ‘occhi’. Per il primo caso (*-ddh*) si è cercato di dare una spiegazione al fenomeno confrontando lo sviluppo *-ll* > *ddh* con simili articolazioni del siciliano, del còrso e dell’italiano meridionale e con suoni enfatici del berbero attribuendolo all’influsso del sostrato paleosardo<sup>54</sup>. Si ipotizza che in termini come *foddha*, *famiddha*, *fiddholu* ci sia stato prima un processo di assimilazione e poi di cacuminazione (attraverso lo schema *-li-* > *-ll-* > *-ddh-*) da *folia*, *familia*, *figliolu(s)*. La cacuminazione si verifica anche:

- con l’affricata mediopalatale sorda *chi* > *chj*: *aricchj*, orecchio, *chjamà*, chiamare, *vècchju*;
- con l’affricata mediopalatale sonora *ghi* > *ghj*: *ghjacciu*, *ghjanda*, *ghjangula*, *ghiandola*;
- con l’affricata prepalatale sonora *ghi* > *ghj* (e *ggi* > *gghj*); *ghjestu*, ‘gesto’, *ghjttà*, ‘gettare’, *ghjocu*, ‘gioco’, *ligghj*, ‘leggere’. Ma la norma non sempre viene osservata: *leggìu*, *reggenza*, *reggioni*;
- con *ga* > *ghja*: *ghjaddina*, ‘gallina’, *ghjaddu*, ‘gallo’;
- in corrispondenza di parole latine terminanti in *-arius* e *-orium*: *capragghju*, *faulagghju*, *limusinagghju* (‘capraio’, ‘bugiardo’, ‘mendicantee’) da *caprarius*, *fabularius*, *elemosinarius*; e così: *cogghju* (lat. *corium*, ‘cuoio’), *rasogghja* (‘coltello a serramanico’, ‘rasoio’). Il fenomeno trova riscontro nel còrso oltremontano, mentre il cismontano ha normalmente le forme di tipo toscano.

- Betacismo: es. còrso *bucértula* (madd. *bicertula*), ‘lucertola’, gall. *bédđula*, ‘donnola’; in Corsica è frequente il passaggio *b* > *v*: bal. *valcone* per *balcone*, oltr. *valzu* accanto a *balzu*, ‘balza’, *vandera* / *bandera* ‘bandiera’. Es. gallurese:

<sup>54</sup> G. MILLARDET, *Sur un ancien substrat commun à la Sicilie, la Corse, la Sardaigne*, in “Revue delinguistique romane”, 9, 1933, pp.346-369.

*abà* (còrso *avali*), ‘ora’, *bacamundu*, ‘vagabondo’, *balariana*, ‘valeriana’, *balisgia*, ‘valigia’, *bibbinnà* (*binninnà*), ‘vendemmiare’, *biddutu*, ‘velluto’, *biscu* (*viscu*), ‘vischio’ (*biscosù*, viscoso), *boci*, ‘voce’ (*bucià*, vociare), *bòitu*, ‘vuoto’ (*sbuità*, ‘vuotare’, ‘svuotare’), *bolu*, ‘volo’ (*vulà*, ‘volare’), *bòmbitu*, ‘vomito’ (*bumbità*, ‘vomitare’), *bulumu* (*vulumu*), ‘volume’, *buscica*, ‘vescica’;

- Formazioni particolari: *ghjàcanu* (‘chierico’, ‘sacrestano’); *ghjalgastolu* (*valgastolu*) (‘gola’) *ghjalmu* (*valmu*) (‘verme’); *ghjastima* e *ghjastimà* (‘bestemmia’ e ‘bestemmiare’); *ghjrabalchinu* (‘menarola’, ‘trapano a mano’) dal piem. *viraberchin*. L’elemento iniziale ‘vira’ (da ‘virare’) si è incrociato con ‘gira’ (gall. *ghjra*) da ‘girare’;
- Comportamenti specifici di alcuni nessi:
  - il nesso *-rn-* si assimila in *-rr-* (assimilazione progressiva): *carri*, ‘carne’, *corru*, ‘corno’, *forru*, ‘forno’, *litarru*, ‘terreno’, *mérula*, ‘merlo’, *turrà*, ‘tornare’; ma *culnicia*, ‘cornice’, *telniu*, ‘terno’, *tolniu*, ‘tornio’;
  - il nesso *-rs-* si assimila in *-ss-* (assimilazione regressiva): *bussa*, ‘borsa’, *Còssica*, ‘Corsica’, *mossu*, ‘morso’ (*mussicà*, ‘morsicare’), *passona*, ‘persona’, *ussu*, ‘orso’, *vassà*, ‘versare’;
  - fenomeni fucali di particolare interesse presentano i nessi *-nj-* e quelli rappresentati dalla *l* e *r* + consonante<sup>55</sup>: la consonante nasale palatale *n*’ (che è scritta comunemente *gn*, ma il digramma è solo grafico, non fonetico) normalmente viene pronunciata appoggiando la lingua alla parte anteriore della cavità orale (palato duro) ma nella pronuncia gallurese la punta della lingua si ritrae verso il velo palatino. Si ha quindi una concomitante sovrapposizione e inversione di suoni, una spinta verso la zona uvulare con incipiente metatesi, -*gn-* > *-ngn-* / *-nghj*.

<sup>55</sup> G. ROHLFS, *Grammatica storica della lingua italiana e dei suoi dialetti*, Torino, Einaudi, 1969, p. 399.

La *r* dinanzi a consonante si risolve in un suono che il Wagner trascrive con *ʎ* e spiega dicendo che «in sostanza la *r* e la cons. un elemento spirante con palatalizzazione della liquida precedente (ba'ka, barca, ful'ka, forca, al'ga, alga, lal'gu, largo)», in aggiunta le palatalizzazioni del tipo *bal'ka* «hanno molta somiglianza con la cosiddetta lisca toscana ed è molto probabile che derivino da questa»<sup>56</sup>.

Ulteriori elementi distintivi del gallurese<sup>57</sup>:

- Infiniti dei verbi: gli infiniti dei verbi sono apocopati come nel còrso e nel sassarese. Gall. *andà, vidè, vindì*; còrso *mancà, vantà* (ma rizotonici quelli in *-e*: *vede*, ‘vedere’, *leghje*, ‘leggere’); sass. *magnà, vidè, suffrì*;
- Complemento oggetto introdotto dalla preposizione *a*: con i verbi transitivi si ha, come nello spagnolo e nel còrso, l'accusativo preposizionale (retto dalla prep. *a*) o personale (solo coi nomi di persona). Gall. *ani briatu a Mariu* (‘hanno rimproverato Mario’); còrso *vogliu vede à ziu Gnaziu* (‘voglio vedere zio Ignazio’);
- Vocativo abbreviato: nei complementi di vocazione di solito si usano le forme accorciate sia coi nomi comuni sia con quelli propri: *O bà, lu femu?* (O ba(bbo), lo facciamo?); *Gjiuà, veni!* (Giova(nni), vieni!);
- *Aidò!*: è un'interiezione caratteristica di tutta la Sardegna e della Corsica. Ha il significato di ‘orsù’, ‘suvvia’. Per il Wagner si tratterebbe di una parola primitiva anetimologica. Si potrebbe avanzare l'ipotesi che essa abbia a che fare col fr. *allons*, ‘orsù’. L'attacco della Marsigliese sembra avere questo significato: “*Allons enfants de la Patrie*” normalmente è tradotto come ‘andiamo, figli della Patria’ e ha in fondo questo significato: ‘Orsù, suvvia, figli della Patria’.

<sup>56</sup> M. L. WAGNER, *Fonetica del sardo* (Introd, trad. e appendice di G. PAULIS), Cagliari, Gianni Trois Editore, 1984, p. 284.

<sup>57</sup> R. DE MARTINO, op. cit., 2001, p. 187.

Anche se si dimostra forte il tratto oltremontano, il gallurese concorda col còrso cismontano per alcuni fenomeni<sup>58</sup>:

- il pronome personale di prima persona (io) il gallurese ha *éu* come i dialetti della valle di Alisani e della zona di Bastia (*eu*) mentre l'oltremontano presenta *éju*, *géju*;
- il gallurese ha *manu* ('mano') come il cismontano (escluso il capocorsino) mentre l'oltremontano ha *mani*;
- per l'assordimento *-ǵ-* > *-č-* della varietà calangianese ha dei riscontri nella fascia mediana che unisce Rennu e Antisanti, mentre l'assordimento *-b-* > *-p-* si trova nelle varietà parlate lungo la linea Vicu – Bocognano – Fiumorbu;
- per la sonorizzazione di *-f-* > *-v-* trova riscontro nel cismontano. Es.: *u vènu* (il fieno) mentre nell'oltremontano prevale l'esito *u βénu*;
- il lessema gallurese *ciòanu* (giovane) ha corrispondenza soltanto nel comune di Cozzano *ciòuanu*, che si trova nel confine linguistico tra cismontano e oltremontano;
- l'esito *kj* > *ts* che si realizza nella varietà di Aggius ha riscontro, oltre che nel bonifacino, nella varietà di Soccia, villaggio che si trova lungo il confine linguistico tra cismontano e oltremontano;
- la sonorizzazione *-k-* > *ǵ* del dialetto di Aggius (es. *lu ǵoddu* 'il collo') è realizzata nelle varietà intermedie di Ciamannaccia, Sampolu, Cozzano e Poggio di Nazza (*u ǵóddu*);
- la sonorizzazione *-p-* > *-b-*. Es. *abbrili* (aprile) ha piena corrispondenza col cismontano mentre l'oltremontano ha *aprili*;

---

<sup>58</sup> M. MAXIA, op. cit., 2008, P. 83.

- il gallurese *paési* (paese) ha riscontro lungo il litorale di Ajaccio e nella fascia mediana che unisce la Cinarca al Fiumorbu, mentre nell'estremo sud prevale *pajési, paésu*;
- il gallurese *stélla* ha corrispondenza nel cismontano mentre l'oltremontano presenta esiti tra *stédde* e *stìdda*.

L'esame del lessico dimostra la prevalenza dell'elemento còrso nel gallurese. Nonostante la presenza di numerose voci sarde queste sono antecedenti all'affermazione della varietà còrsa, quindi possono essere considerate come un sostrato. Inoltre la componente lessicale sarda di frequente presenta sinonimi còrsi e adeguamenti fonetici. In aggiunta nel gallurese sono assenti molte voci che invece contraddistinguono il sardo: *chenabura*, *domo*, *berbeghe* o *inoghe*, di fronte al tipo còrso e italiano *vennari* (venerdì), *casa* (casa), *pecura* (pecora) e *chici* (qui)<sup>59</sup>.

Proseguendo nell'analisi del lessico si può asserire che i genovesismi siano penetrati in Gallura attraverso il còrso, il sassarese o a causa dei contatti commerciali con l'isola linguistica di Bonifacio e con la Liguria. Sono però assenti molti ligurismi che si trovano nel còrso, dato che conferma la stabilità dell'impianto còrso in Sardegna con largo anticipo rispetto alla tradizionale cronologia settecentesca<sup>60</sup>.

Attraverso l'esame del lessico appare evidente il legame del gallurese col còrso meridionale pur di fronte a un influsso sardo. La tempistica degli insediamenti còrsi esclude definitivamente un significativo influsso toscano nella formazione del gallurese e ha sempre meno senso parlare di caratteri continentali del gallurese, difatti quello che accomuna questa varietà ai dialetti toscani appare costantemente condiviso, nel lessico come per gli altri aspetti, anche dal còrso<sup>61</sup>.

---

<sup>59</sup> F. TOSO, op. cit., 2012, P. 32 e segg.

<sup>60</sup> Ibid.

<sup>61</sup> Ibid.



## I. 4 La cultura gallurese

Il nord Sardegna e la parte meridionale della Corsica sono due terre sorelle inestricabilmente legate dalla stessa storia e dallo stesso retaggio culturale e linguistico. Per tale ragione galluresi e sassaresi non si sentono sardi. Se i galluresi chiamano gli altri sardi *li Saldi* (i Sardi), con una serie di sfumature determinate dalla lontananza delle varie aree dialettali della Gallura, e se i sassaresi chiamano i logudoresofoni *li di li biddi* (quelli dei paesi), si può affermare che essi non si sentano sardi e percepiscano la loro condizione di alloglossia. Per di più sia i sassaresi sia i galluresi definiscono i Corsi *li Còssi* o *li Cossicani* e non percepiscono un senso di appartenenza nemmeno a tale etnia. Specialmente nei galluresi si coglie un sentimento di appartenenza a una microetnia diverso rispetto all'ampio concetto di sardità; i maddalenini invece avvertono un'appartenenza caratterizzata più in senso còrso che sardo<sup>62</sup>.

Di conseguenza quella del popolo gallurese è una cultura diversa da quella del resto della Sardegna e può essere compresa sotto il nome di Cultura degli Stazzi. La Gallura degli stazzi ha origini antiche ma ha assunto la forma che noi conosciamo oggi negli ultimi secoli. La sua originalità è data da un insediamento detto ad habitat disperso: piccoli nuclei familiari diffusi in modo rado nel territorio<sup>63</sup>.

L'essenza della civiltà gallurese affonda le sue radici nel modello produttivo basato sull'economia dello stazzo. Gli autori di questa forma di stanzialità sono pastori e allo stesso tempo contadini fieri e orgogliosi. Questi erano possidenti soddisfatti della loro condizione economica e avevano una rosea visione del futuro; costruivano grandi proprietà e possedevano vasti appezzamenti di terreni coltivabili; conoscevano il mestiere dell'agricoltore ed erano in grado di far fronte alle calamità naturali; vivevano in libertà lontani da qualsiasi forma di pressione sociale. Anche le tasse e le impostazioni feudali erano dimenticate e ciò esaltava la sensazione di autonomia del pastore-contadino rafforzando la sua capacità organizzativa<sup>64</sup>.

Una volta affermato il ruolo di pastore-contadino lo stazzo mutò di significato. Non indicava più solo un terreno ma la casa del pastore e della sua famiglia: vi trovavano posto l'orto, la vigna, il frutteto, i locali per la produzione del formaggio e dei salami, il forno per il pane, strumenti per filare la lana e il telaio. Era un mondo

---

<sup>62</sup> M. MAXIA, op. cit., 2008, p.201.

<sup>63</sup> A cura di CONSULTA INTERCOMUNALE GALLURESE, op. cit., 2003, pp. 10-11.

<sup>64</sup> Ivi, p.13.

pastorale autosufficiente in cui ogni membro della famiglia aveva un preciso ruolo produttivo<sup>65</sup>. La coltura dei cereali soddisfaceva le esigenze della ristretta comunità che poco risentiva delle variazioni generali dell'economia, a esclusione dei periodi di carestia. La casa rurale era quindi un simbolo di relativo benessere e nei periodi di crisi la Gallura con le sue risorse differenziate nonostante la povertà dei terreni e le difficoltà delle comunicazioni, era una delle regioni meno povere dell'isola<sup>66</sup>.

I campi aperti venivano chiamati *vidazzone*: al centro di essi stava la villa con spazi recintati limitati per colture speciali; le terre a riposo dopo la coltivazione e poi aperte al pascolo comune degli animali erano dette *paberile*; i terreni demaniali sottoposti al controllo delle comunità erano i *salti* e per il loro utilizzo si pagava un tributo chiamato *ademprivio*. Chi seminava nei *salti* non poteva beneficiare della protezione del villaggio ed era esposto agli assalti, ai danneggiamenti e alle devastazioni. Tuttavia in Gallura, l'appropriazione rischiosa, il popolamento diffuso, il radicamento di nuclei e famiglie contadine in un territorio seppur incolto e inospitale fu comunque costante e favorito dalla speranza di una vita migliore e lontana dai villaggi popolosi e statici<sup>67</sup>.

La formazione degli stazzi ha anche un'altra caratteristica originale: la colonizzazione còrsa. Numerose famiglie di pastori bonifacini praticavano la transumanza sulle coste sarde e sulle vicine isole minori. Molti erano i contrabbandieri che, sostenuti dai pastori locali, si rifugiavano in Gallura per sfuggire alle vendette tra famiglie e alle crisi sociali. Tale migrazione fece aumentare il numero degli stazzi: agli inizi del XVIII secolo coloro che parlavano un dialetto còrso, seppure sardizzato, erano la maggioranza della popolazione negli stessi villaggi di Tempio e satelliti e la quasi totalità negli stazzi. Resistevano poche isole logudoresi: Luras, Olbia e Bortigiadas. Questo ha fatto della Gallura una particolarità etnica e linguistica<sup>68</sup>.

La Gallura agli inizi del XIX sec. aveva quindi già le odierne sembianze e gli odierni costumi che non si sono ancora persi<sup>69</sup>. Le feste campestri avevano una funzione molto importante perché erano espressione di tutto ciò che rappresentava la vita nello stazzo e testimoniavano un modello culturale nel quale si riconosceva l'identità del popolo gallurese; costituivano inoltre un momento di socializzazione dove trovare aiuto

---

<sup>65</sup> A cura di CONSULTA INTERCOMUNALE GALLURESE, op. cit., 2003, p. 11.

<sup>66</sup> T. PANU, op. cit., 2010, p. 185.

<sup>67</sup> A cura di CONSULTA INTERCOMUNALE GALLURESE, op. cit., 2003, p. 12 e segg.

<sup>68</sup> Ibid.

<sup>69</sup> A cura di CONSULTA INTERCOMUNALE GALLURESE, op. cit., 2003, p. 14.

reciproco e protezione da ciò che si poteva percepire come pericolo<sup>70</sup>. Canti e balli rappresentavano sia un codice di espressione comunitaria sia l'apertura e il consolidamento di quei valori che caratterizzavano il vivere quotidiano. Si presuppone che il fenomeno delle chiese campestri diffuse affondi le sue radici fino a epoche prenuragiche<sup>71</sup>.

Quando le ville vennero distrutte e abbandonate a causa dei continui assalti e tentativi di invasione da parte di saraceni e barbari, le chiese resistettero in rispetto del sentimento religioso. Attorno a esse, fino ai secoli XVI-XVII-XVIII, continuarono a ripetersi in forma di festa in onore del santo, gli incontri comunitari di popolazioni ad habitat disperso e degli abitanti delle villae residue<sup>72</sup>. Oggi le feste campestri costituiscono ancora un elemento originale di vivacità che in qualche modo salvaguarda le tradizioni di fronte ai profondi cambiamenti determinati dalla modernità e «mantenendo un equilibrio precario tra il vecchio e il nuovo, potrebbero spingere a capire meglio i bisogni che vi si esprimono ed a comprendere le ragioni collettive della gente»<sup>73</sup>. In esse si trovano ancora i codici comunicativi, linguistici e gestuali e gastronomici che definiscono la specificità gallurese proprio per la grande e inalterata partecipazione popolare che ripete le azioni quotidiane del passato in forma cerimoniale.

Sono stati finora elencati i fondamenti costitutivi della cultura gallurese che si riflettono nella letteratura popolare.

Il già citato Gavino Pes, noto anche come Don Baignu, è l'iniziatore della poesia gallurese di tipo colto e si faceva promotore della rappresentazione della vita quotidiana delle genti galluresi. Le opere di Pes e la sua poesia popolare sono percepite come fattore costitutivo dell'identità gallurese e si inserirono rapidamente nella tradizione orale. Alcuni verseggiatori popolari, imitatori a un livello più basso dell'opera di Pes fecero sì che i suoi versi fungessero da «collegamento tra un nucleo abitato e l'altro, e come strumento di individuazione e fissazione di una "civiltà gallurese" »<sup>74</sup>.

Don Baignu fu anche capo di una piccola Arcadia gallurese, e le sue opere rappresentano un modello di vita forse «privo di grandi idealità e di virtù, senza drammi e senza esaltazioni: un mondo di sobrietà e di semplicità, di rapporti umani schietti e

---

<sup>70</sup> M. ATZORI, M. M. SATTÀ, *Credenze e riti magici in Sardegna*, Sassari, Chiarella, 1980, pp. 175-192.

<sup>71</sup> G. BELLU, rel. CLARA GALLINI, tesi di laurea: *Storia religiosa gallurese: alcuni aspetti*, Università di Cagliari, a. a. 1970.

<sup>72</sup> R. TURTAS, *Le chiese nel verde*, Sassari, 1988.

<sup>73</sup> G. DEIDDA, A. DELLA MARIA, *Sagre, riti e feste popolari in Sardegna*, Cagliari, Janus, 1987.

<sup>74</sup> TOLA 2006: 239, in F. TOSO, op. cit. 2012, p. 37.

precisi» che corrisponde a «quel mondo contadino generalmente gentile ed educato che è il mondo della civiltà gallurese»<sup>75</sup>.

La linea di poesia alta ha conosciuto un grande sviluppo nel corso dell'ultimo secolo grazie all'opera di alcuni autori tempiesi: Giulio Cossu (Tempio 1920-2014) rappresenta il prototipo del poeta colto, insegnante appassionato delle tradizioni linguistiche e culturali locali ma anche conferenziere e organizzatore di iniziative per la valorizzazione del gallurese (per il quale ha elaborato una prima codificazione ortografica sulla quale si basano sostanzialmente gli usi scritti attuali)<sup>76</sup>.

Il gallurese vanta anche una discreta tradizione come lingua del teatro (MUZZEDDU 2010; ACHENZA 2009) e della canzone corale o individuale: Fabrizio De Andrè usò il gallurese in due canzoni perfettamente inserite nell'opera di ricerca e d'innovazione sonora svolta dal cantautore genovese che risedette a lungo nei pressi di Tempio<sup>77</sup>.

Recentemente nel nord dell'isola sono sorte alcune iniziative di difesa e valorizzazione delle parlate di origine còrsa. I galluresi dimostrano di essere i più determinati: questi da una decina d'anni hanno realizzato una Consulta Intercomunale Gallurese con sede ad Arzachena e l'Accademia *di la linga gadduresa* a Luogosanto. La prima è composta dai sindaci dei comuni corsòfoni e da alcuni intellettuali e sembra svolgere un ruolo sia politico sia linguistico, la seconda ha finalità prettamente linguistiche e si compone di docenti universitari, dirigenti scolastici, insegnanti e operatori culturali: obiettivo comune è la tutela e la promozione del gallurese attraverso lo studio e la pubblicazione di saggi e volumi<sup>78</sup>.

Il canto e il ballo tradizionali contribuiscono al perfetto inserimento di gruppi folkloristici galluresi in un vivido panorama di manifestazioni culturali. Queste iniziative animano trasmissioni radiofoniche e televisive della Sardegna e rappresentano uno degli aspetti di maggior coesione sociale dell'isola. I programmi televisivi *Sardegna canta*, dell'emittente Videolina di Cagliari, e *Buonasera Sardegna* dell'altra emittente cagliaritana Sardegna Uno, sono i programmi locali di intrattenimento più seguiti nell'isola. La trasmissione radiofonica arzachenese, *La boci di la Gaddura* (La voce della Gallura) dai contenuti popolari, al presente è la più seguita e trasmette

---

<sup>75</sup> BRIGAGLIA 1975:84 in F. TOSO, op. cit. 2012, p. 38.

<sup>76</sup> F. TOSO, op. cit. 2012, p.38.

<sup>77</sup> Ibid.

<sup>78</sup> F. TOSO, op. cit. 2012, p. 39.

canzoni dell'intero repertorio tradizionale sardo. Questo programma radiofonico, pur essendo condotto interamente in gallurese viene ascoltato anche dai sardofoni del Logudoro e del Campidano che spesso interloquiscono con i conduttori nelle rispettive varietà<sup>79</sup>.

Il gallurese ambisce a perseverare la propria parlata dalla decadenza che minaccia tutte le lingue minori. La perdita della lingua inoltre porterebbe alla perdita di quei caratteri che fanno della civiltà gallurese un esempio unico di commistione di varie influenze che rende l'unicità culturale della Gallura e dei galluresi un patrimonio inestimabile<sup>80</sup>.

Il buon gallurese nasce dal superamento del contrasto tra pastore e contadino che ha cancellato gli atteggiamenti violenti presenti ancora nei territori confinanti, questo sta alla base dei tratti più gentili della Gallura. Il pastore-contadino gallurese è stato prima possessore e poi proprietario del suo fondo: si è ribellato alle servitù esterne ed è stato capace di esprimersi con grande spirito d'iniziativa<sup>81</sup>.

L'isolamento della famiglia pastorale nello stazzo è a fondamento dell'individualismo tipicamente gallurese che si accompagna a un senso di forte autostima e autonomia. Nonostante il grande valore attribuito a tale individualità, la solidarietà sociale era ed è ancora grande<sup>82</sup>.

Altro fondamento che istituisce la civiltà gallurese è la garanzia reciproca, il riconoscimento e la volontà di permettere a ognuno di prosperare nel proprio benessere. Si tratta di una dimensione differente sebbene le origini comuni, rispetto alla solitudine del pastore barbaricino che lotta contro tutti per la sopravvivenza del suo gregge. Quindi quella gallurese è una civiltà dai tratti unici che riconosce pari diritti e dignità personale di ognuno dei suoi componenti. Questi

fondamenti materiali fanno essere i Galluresi indulgenti tra loro; coltivano tratti e modi cortesi, hanno la capacità di risolvere nei modi civili anche i contrasti più vivaci. Qualità tutte che sono riconosciute universalmente alla gente gallurese. Quella cultura può essere feconda anche nel riconoscimento delle ragioni degli altri e nel superamento delle diversità<sup>83</sup>.

---

<sup>79</sup> M. MAXIA, op. cit., 2008, p. 219.

<sup>80</sup> A cura di CONSULTA INTERCOMUNALE GALLURESE, op. cit., 2003, pp. 16-17.

<sup>81</sup> Ivi, pp. 18-19.

<sup>82</sup> Ibid.

<sup>83</sup> A cura di CONSULTA INTERCOMUNALE GALLURESE, op. cit., 2003, p. 11.

Queste rappresentano le antiche radici da cui è scaturito lo stimolo di proteggere quei valori che oggi vanno affievolendosi, fino a far temere la loro definitiva scomparsa. Nella situazione attuale sarebbe opportuno rinforzare le fondamenta culturali galluresi tessendo una nuova trama di solidarietà che faccia tesoro dei risultati e dello spirito espresso dalla cultura degli stazzi. Senza la salvaguardia del suo codice linguistico e della conoscenza intrinseca a quel codice, la cultura degli stazzi è destinata a scomparire.

## Capitolo II

### La tutela delle minoranze linguistiche

Le lingue sono un elemento di grande ricchezza nel patrimonio culturale di un popolo: costituiscono infatti uno dei più importanti sviluppi del genere umano, se non addirittura il più importante poiché, oltre la prospettiva della comunicazione, portano con sé un bagaglio di valori e di credenze. Le lingue esprimono l'identità dei loro parlanti e delineano il loro passato, il loro presente e il loro futuro: custodiscono radicata in sé una grande conoscenza del mondo e dell'esperienza umana. La lingua che parliamo contribuisce infatti a determinare chi siamo<sup>1</sup>. Date queste premesse risulta quasi scontato affermare che nel momento in cui una lingua muore, il retaggio relativo a essa viene perduto. Nel corso degli ultimi cinquant'anni, ad esempio, il processo di globalizzazione ha diffuso a livello mondiale una cultura anch'essa globale di cui la lingua inglese è il motore trainante; molte lingue meno utilizzate hanno difficoltà a competere e sopravvivere in un simile contesto. Insieme all'inglese, poche lingue di grande prestigio dominano i media e il mercato, il sistema scolastico e burocratico, così le culture locali, più deboli, vengono messe da parte<sup>2</sup>.

Negli anni Novanta l'UNESCO ha pubblicato il *Red Book of Endangered Languages*, un elenco delle lingue a rischio di estinzione nel mondo, poi sostituito dall'*Atlas of the World's Languages in Danger*<sup>3</sup>. Nel 2002-2003, la stessa UNESCO ha chiesto a un gruppo internazionale di esperti di elaborare un quadro specifico per classificare la vitalità di una lingua. Il quadro definisce sette livelli di vitalità, cinque dei quali descrivono la condizione delle lingue a rischio di estinzione: sicura, sicura ma minacciata, vulnerabile, certamente in pericolo, in grave pericolo, criticamente in pericolo ed estinta<sup>4</sup>. Le lingue in pericolo nell'Unione Europea sono 128 e la minoranza gallurese risulta tra quelle certamente in pericolo. «Una lingua è in pericolo quando i parlanti smettono di usarla, se ne servono in un numero sempre minore di contesti comunicativi e smettono di trasmetterla di generazione in generazione. Ciò significa che

---

<sup>1</sup> COMMISSIONE EUROPEA, *Parlare le lingue europee, Le lingue nell'Unione europea*, ec.europa.eu/publications, Direzione generale della Comunicazione, Pubblicazioni, Lussemburgo, Ufficio delle pubblicazioni ufficiali delle Comunità europee, 2008, pag.3.

<sup>2</sup> M.P. JONES, *Lingue a rischio di estinzione e diversità linguistica nell'Unione Europea*, Parlamento Europeo, 2013, p.4.

<sup>3</sup> <http://www.unesco.org/languages-atlas/>

<sup>4</sup> Ibid.

vengono a mancare nuovi parlanti, siano questi adulti o bambini»<sup>5</sup>. Il desiderio di tramandare una lingua minoritaria dipende quindi dall'esercizio della volontà popolare: spesso quando quest'ultima viene a mancare, le misure di tutela vengono percepite alla stregua di un hobby per pochi cultori, studiosi ed esperti<sup>6</sup>. Un cambiamento di percezione risiede in quello che rappresenta il valore di questa scelta, attuabile modernizzando il concetto stesso di tutela ed elaborando nuove strategie. I fenomeni di bilinguismo e di multilinguismo potrebbero costituire infatti una grande risorsa anche in termini di creatività e di innovazione<sup>7</sup>.

---

<sup>5</sup> UNESCO, 2003.

<sup>6</sup> M. ABLEY, *Spoken Here*, Berkshire (UK), Arrow Books, 2003, p.4.

<sup>7</sup> M.P. JONES, *Lingue a rischio di estinzione e diversità linguistica nell'Unione Europea*, Parlamento Europeo, 2013, p.4.



## II. 1 La normativa europea

Il rispetto della diversità linguistica e culturale costituisce un tratto distintivo dell'Unione europea (UE)<sup>8</sup>. L'UE riconosce il diritto all'identità e promuove attivamente la libertà di parlare e scrivere nella propria lingua continuando comunque a perseguire l'obiettivo di una stretta integrazione tra i paesi membri. In tali obiettivi si incarna il motto dell'UE, la quale si dichiara «unita nella diversità»<sup>9</sup>. Tale principio viene anche sancito dalla *Carta europea dei diritti fondamentali* adottata dai leader dell'UE nel 2000<sup>10</sup>. Le lingue regionali o minoritarie sono invece tutelate più nello specifico da una carta approvata nell'ambito del Consiglio d'Europa.

### II. 1. 1 La Carta europea delle lingue regionali o minoritarie

La *Carta europea delle lingue regionali o minoritarie* è un trattato europeo adottato a Strasburgo il 5 novembre 1992 sotto l'egida del Consiglio d'Europa per proteggere e promuovere le lingue regionali e minoritarie storiche in Europa<sup>11</sup>. Essa rappresenta il testo guida cui conformare le diverse leggi nazionali in materia di tutela delle minoranze linguistiche. Lo scopo della Carta è quello di contribuire a conservare e a sviluppare le tradizioni e la ricchezza culturali d'Europa, nonché realizzare una stretta congiunzione fra i suoi membri, specialmente al fine di salvaguardare e di promuovere gli ideali e i principi che sono il loro patrimonio comune. Il provvedimento è stato recepito solo da alcuni dei paesi membri del Consiglio d'Europa: diciotto paesi europei hanno firmato e ratificato la Carta. L'Italia ha sottoscritto la Carta il 27 giugno 2000 ma deve ancora provvedere alla sua ratifica legislativa.

Ai sensi della Carta con l'espressione «lingue regionali o minoritarie» si intendono le lingue «usate tradizionalmente sul territorio di uno Stato dai cittadini di detto Stato che formano un gruppo numericamente inferiore al resto della popolazione

---

<sup>8</sup> COMMISSIONE EUROPEA, *Parlare le lingue europee, Le lingue nell'Unione europea*, ec.europa.eu/publications, Direzione generale della Comunicazione, Pubblicazioni, Lussemburgo, Ufficio delle pubblicazioni ufficiali delle Comunità europee, 2008, pag. 7.

<sup>9</sup> COMMISSIONE EUROPEA, *Parlare le lingue europee, Le lingue nell'Unione europea*, ec.europa.eu/publications, Direzione generale della Comunicazione, Pubblicazioni, Lussemburgo, Ufficio delle pubblicazioni ufficiali delle Comunità europee, 2008, pag.3.

<sup>10</sup> EUROPEA, Unione. *Carta dei diritti fondamentali dell'Unione Europea*, disponibile al sito: [www.europarl.europa.eu/charter/pdf/text\\_it.pdf](http://www.europarl.europa.eu/charter/pdf/text_it.pdf), 2000.

<sup>11</sup> *Carta europea delle lingue regionali o minoritarie*, Consiglio d'Europa, Preambolo, 1992.

dello Stato; e diverse dalla (e) lingua (e) ufficiale (i) di detto Stato»<sup>12</sup>. Generalmente si riconoscono tre categorie di lingue regionali e minoritarie<sup>13</sup>: 1. lingue specifiche per una regione che può trovarsi all'interno di uno Stato membro o regione transfrontaliera e che non costituiscono una lingua dominante in nessun paese dell'Unione. Tra queste vi sono lingue quali il basco, il bretone, il catalano, il frisone, il sardo e il gallese; 2. lingue parlate da una minoranza della popolazione in un paese UE ma che sono lingue ufficiali in un altro: per esempio il tedesco nel sud della Danimarca, il francese in Valle d'Aosta, nel nord dell'Italia e l'ungherese in Slovacchia; 3. lingue non territoriali quali quelle delle comunità rom o ebraiche nell'UE (romani e yiddih). La Carta inoltre specifica che «questa espressione (lingue regionali o minoritarie) non include né i dialetti della(e) lingua(e) ufficiale(i) dello Stato né le lingue dei migranti»<sup>14</sup>. È facilmente deducibile quindi che le varietà alloglotte sarde e nello specifico la variante gallurese sono escluse da tale tipo di tutela. Per «territorio» la Carta intende «l'area geografica nella quale tale lingua è l'espressione di un numero di persone tale da giustificare l'adozione di differenti misure di protezione e di promozione»<sup>15</sup>. Per «lingue non territoriali» intende invece «le lingue usate da alcuni cittadini dello Stato che differiscono dalla(e) lingua(e) usata (e) dal resto della popolazione di detto Stato ma che, sebbene siano usate tradizionalmente sul territorio dello Stato, non possono essere ricollegate a un'area geografica particolare di quest'ultimo»<sup>16</sup>.

Gli obiettivi e i principi perseguiti dalla Carta sono<sup>17</sup>:

- il riconoscimento delle lingue regionali o minoritarie come espressione della ricchezza culturale;
- il rispetto dell'area geografica di ogni lingua regionale o minoritaria per fare in modo che le divisioni amministrative esistenti o nuove non costituiscano un ostacolo alla promozione di una lingua regionale o minoritaria;

---

<sup>12</sup> *Carta europea delle lingue regionali o minoritarie*, Consiglio d'Europa, Art. 1, Parte I, Disposizioni generali, Definizioni.

<sup>13</sup> COMMISSIONE EUROPEA, *Parlare le lingue europee, Le lingue nell'Unione europea*, ec.europa.eu/publications, Direzione generale della Comunicazione, Pubblicazioni, Lussemburgo, Ufficio delle pubblicazioni ufficiali delle Comunità europee, 2008, p.7

<sup>14</sup> *Carta europea delle lingue regionali o minoritarie*, Consiglio d'Europa, 1992, Preambolo.

<sup>15</sup> *Ibid.*

<sup>16</sup> *Ibid.*

<sup>17</sup> *Ivi*, 1992, Art. 7, Obiettivi e principi.

- il riconoscimento della necessità di un'azione risoluta di promozione delle lingue regionali o minoritarie allo scopo di salvaguardarle;
- la facilitazione e/o l'incoraggiamento dell'uso orale e scritto delle lingue regionali o minoritarie nella vita pubblica e nella vita privata;
- il mantenimento e lo sviluppo di relazioni, negli ambiti contemplati dalla Carta, fra i gruppi che usano una lingua regionale o minoritaria e altri gruppi dello stesso Stato che parlano una lingua usata in una forma identica o simile, come pure l'instaurarsi di relazioni culturali con altri gruppi dello Stato che usano lingue diverse;
- la messa a disposizione di mezzi adeguati di insegnamento e di studio delle lingue regionali o minoritarie che consentano ai non parlanti che abitano nell'area in cui tale lingua è usata, di apprendere se essi lo desiderano;
- la promozione di studi e ricerche sulle lingue regionali o minoritarie nelle università o in istituti equivalenti;
- la promozione delle forme appropriate di scambi sopranazionali, negli ambiti contemplati dalla Carta, per le lingue regionali o minoritarie usate in forma identica o simile in due o più Stati.

Il Comitato di esperti che coadiuva i lavori del Consiglio d'Europa ha comunque osservato che in molti Stati manca ancora un approccio strutturato alla conservazione e alla promozione delle lingue: per questo motivo raccomanda che gli Stati firmatari elaborino strategie di lungo periodo per proteggere le lingue minoritarie<sup>18</sup>.

---

<sup>18</sup> *Carta europea delle lingue regionali o minoritarie*, Consiglio d'Europa, Art. 7, Obiettivi e principi.

## II. 1. 2 La Convenzione-quadro per la protezione delle minoranze nazionali

L'evoluzione storica europea dimostra che la protezione delle minoranze nazionali è necessaria per la stabilità, la sicurezza democratica e la pace del continente. Una società pluralistica e democratica deve infatti rispettare l'identità etnica, culturale, linguistica e religiosa di ogni persona appartenente a una minoranza nazionale, e creare le condizioni adatte a permettere a data minoranza di esprimere, preservare e sviluppare questa identità. In questa prospettiva l'UE nel 1995 ha adottato un ulteriore strumento di tutela delle minoranze: *la Convenzione-quadro per la protezione delle minoranze nazionali* (entrata in vigore il 01/02/1998)<sup>19</sup>.

Obiettivo della Convenzione è la promozione di condizioni adatte a permettere alle persone appartenenti a minoranze nazionali di conservare e sviluppare la propria cultura e preservare i più importanti elementi identitari, cioè la loro religione, la loro lingua, le loro tradizioni così come il loro patrimonio culturale. La Convenzione-quadro non contiene una definizione del concetto di «minoranza nazionale» poiché non esiste una definizione per la quale concordano tutti gli Stati membri del Consiglio d'Europa; viene così lasciata la possibilità a ogni Stato contraente di precisare a quali gruppi minoritari presenti sul proprio territorio debbano essere applicate le disposizioni della Convenzione.

Tutte le esperienze europee relative alle minoranze linguistiche e nazionali hanno messo in rilievo che le azioni di rivitalizzazione di una minoranza non possono essere confuse o intrecciate con quelle della cultura generale o globale della popolazione, della regione o dello Stato di cui la minoranza fa parte. La cultura intesa come salvaguardia del patrimonio materiale e immateriale è altra cosa rispetto alla politica linguistica: i problemi di pianificazione e rinascenza linguistica necessitano di una legislazione e di un'attività dedicata. Il ruolo attribuito alle realtà minoritarie si sta avviando verso un'evoluzione positiva anche in ragione della minore portata eversiva attribuita ai processi rivendicativi connessi. La lingua di minoranza comunque è ancora percepita dalla popolazione come un insieme di dialetti inferiori e minori, e ha bisogno di essere sostenuta, vincolata e protetta<sup>20</sup>.

---

<sup>19</sup> <http://www.coe.int/en/web/conventions/full-list/-/conventions/treaty/157>

<sup>20</sup> Relazione di accompagnamento al D.D.L. in materia di “Norme per la tutela, valorizzazione e promozione della lingua sarda e delle altre varietà linguistiche della Sardegna”, Regione Autonoma della Sardegna, 1997, p.10.

Il percorso di tutela trova comunque l'ostacolo della definizione del concetto di «minoranza linguistica» che varia a seconda della nazione e della situazione linguistica, tanto da condizionare la prassi giuridica e i provvedimenti legislativi. L'esigenza della tutela e di promozione dei patrimoni linguistici regionali e minoritari è ormai ampiamente accolta a livello comunitario, ma continuano a mancare indicazioni generali su cosa si debba intendere per tutela e promozione. Linguisti e giuristi sembrano concordare sul fatto che un'effettiva tutela dei patrimoni linguistici debba articolarsi su due linee d'intervento: occorre tenere distinto il problema della tutela dei diritti linguistici dei parlanti (ossia la non discriminazione di essi in base alla lingua) da quello della tutela e della promozione del patrimonio linguistico, soprattutto di quello tradizionale, in quanto bene culturale ed elemento costitutivo dell'identità comune nazionale ed europea<sup>21</sup>. Alla tutela dei patrimoni linguistici tradizionali in quanto bene culturale dovrebbe corrispondere una tutela integrale di tale bene, nella quale vanno incluse tutte le forme tradizionali di espressione: nel caso italiano, ad esempio, la lingua nazionale, i dialetti italiani, le lingue e i dialetti di tutte le minoranze linguistiche, i gerghi, le lingue liturgiche di alcuni gruppi religiosi e anche le lingue degli immigrati o le lingue straniere diffuse. È evidente che le forme di difesa di tale bene, affidate anche alla responsabilità individuale dei parlanti, richiederebbero l'affermarsi di una sensibilità diffusa in merito al problema che preveda il coinvolgimento di agenzie culturali ed educative parallelamente all'intervento legislativo<sup>22</sup>.

---

<sup>21</sup> F. TOSO, (2009) *Patrimoni linguistici e lingue minoritarie: la prospettiva europea e quella italiana*, "Annali della Facoltà di Lingue e Letterature Straniere dell'Università di Sassari", Vol. 5 (2005 pubbl. 2009), p. 115-124.

<sup>22</sup> F. TOSO, *Il Patrimonio linguistico sardo* Parte VIII – Aspetti del panorama linguistico europeo, 20. LINGUE E DIALETTI.

## II. 2 La normativa italiana

L'Italia pur non avendo ratificato la *Carta europea delle lingue regionali o minoritarie* ha comunque assicurato una prima attuazione delle sue indicazioni attraverso la L. N. 15 dicembre 1999 n. 482 *Norme in materia di tutela delle minoranze linguistiche storiche*. In attuazione del programma di Governo l'Amministrazione della Regione Sardegna, richiamandosi alla normativa europea in materia di minoranze linguistiche, e in particolare all'art. 6 della Costituzione, alla *Carta europea per le lingue regionali e minoritarie* e in attuazione dei principi contenuti nella L. R. 15 ottobre 1997 n. 26 in materia di *Promozione e valorizzazione della cultura e della lingua della Sardegna* ha realizzato iniziative volte a sostenere il ruolo della lingua sarda e delle varietà linguistiche alloglotte presenti nel territorio.

### II. 2. 1 L. R. 15 ottobre 1997 n. 26

Il D. D. L. in materia di *Norme per la tutela, valorizzazione e promozione della lingua sarda e delle altre varietà linguistiche della Sardegna* costituisce uno degli atti principali di Governo attraverso il quale si realizza il progetto identitario per la Sardegna, inteso quale processo sempre in via di elaborazione che necessita di un forte radicamento nella conoscenza, nella memoria della comunità e nelle capacità di progettare un futuro di autonomia comunitaria sostanziale e responsabile<sup>23</sup>.

Ai sensi della presente legge la Regione assume come beni fondamentali da valorizzare la lingua sarda – riconoscendole pari dignità rispetto alla lingua italiana – la storia, le tradizioni di vita e di lavoro, la produzione letteraria scritta e orale, l'espressione artistica e musicale, la ricerca tecnica e scientifica, il patrimonio culturale del popolo sardo nella sua specificità e originalità, nei suoi aspetti materiali e spirituali<sup>24</sup>.

Come esplicito nella relazione di accompagnamento alla legge, per «lingua sarda» si è voluto intendere

---

<sup>23</sup> Relazione di accompagnamento al D.D.L. in materia di “Norme per la tutela, valorizzazione e promozione della lingua sarda e delle altre varietà linguistiche della Sardegna”, Regione Autonoma della Sardegna, 1997, p. 1.

<sup>24</sup> Legge regionale 15 ottobre 97 numero 26, Titolo I (principi e finalità), articolo due oggetto, punto 1, “promozione e valorizzazione della cultura e della lingua della Sardegna.

così come già sottolineato e messo in evidenza da studi approfonditi e riconosciuti dalla comunità scientifica internazionale e asseverati dal legislatore sia statale che regionale, quel sistema-famiglia di varietà linguistiche evolutesi nell'Isola a partire da un sostrato latino originario che pur nella naturale variazione, mantengono comunque una serie di tratti originali e tipici nell'evoluzione dal latino in volgare. La lingua sarda è comunque parlata storicamente nella gran parte del territorio regionale. Il sardo, lingua di identità storia e propria della Sardegna, all'interno del suo dominio si suddivide in numerose varietà territoriali e locali (oltre a subire tutti i mutamenti costanti a cui sono soggette normalmente le varietà linguistiche non standardizzate dal punto di vista cronologico, sociologico, personale, mediale, funzionale) che variano nelle classificazioni convenzionali operate da linguisti e dialettologi fino a singoli esiti, come in ogni lingua naturale, in ogni municipio della Sardegna<sup>25</sup>.

[...]La medesima valenza attribuita alla cultura ed alla lingua sarda è riconosciuta con riferimento al territorio interessato, alla cultura ed alla lingua catalana di Alghero, al tabarchino delle isole del Sulcis, al dialetto sassarese e a quello gallurese<sup>26</sup>.

La questione del rapporto tra la lingua sarda e le varietà alloglotte dell'Isola è centrale nella definizione del testo di legge e si caratterizza per l'ipotetica parità di condizioni e interventi nel riconoscimento reciproco distintivo dal punto di vista identitario e culturale. I termini lingua e dialetto sono fondamentalmente sinonimi per quanto riguarda l'oggetto che definiscono, ma implicano importanti sfumature rispetto ai ruoli e alle attribuzioni che l'oggetto stesso di volta in volta assume. Lingua e dialetto sono sistemi di elementi fonetici, grammaticali e lessicali articolati per la comunicazione orale e scritta. Il dialetto non è quindi, per sua natura, qualcosa di subordinato rispetto alla lingua: ha identiche possibilità espressive, e condivide con la lingua la capacità di arricchirsi, aggiornarsi e rinnovarsi per allargare il proprio ruolo nella trasmissione e la definizione dei concetti<sup>27</sup>.

La definizione di dialetto sembra proporsi soprattutto in negativo, come strumento di comunicazione non formalizzato, subordinato, non istituzionalizzato. Tale definizione ha valore soltanto in contrapposizione al concetto di lingua. I dialetti generalmente mancano di terminologia scientifica se comparati a lingue come l'inglese o il giapponese, alcuni non hanno una letteratura scritta, ma la capacità di dare forma all'esperienza esiste in tutti i linguaggi umani: alcuni idiomi hanno un vocabolario più ampio rispetto ad altri, altri dispongono di una grande varietà di suoni, ma nessuna lingua è migliore di un'altra; le lingue sono al di sopra delle gerarchie. Nemmeno la limitatezza geografica dell'area in cui un idioma è parlato o il ristretto numero dei

---

<sup>25</sup> Regione Autonoma della Sardegna – Relazione di accompagnamento al disegno di legge “Norme per la tutela, valorizzazione e promozione della lingua sarda e delle altre varietà linguistiche della Sardegna”. L. 15 ottobre 97 numero 26, p. 4.

<sup>26</sup> Titolo I (principi e finalità) Art. 2 oggetto della 26, punto 4 Titolo I (principi e finalità) Legge regionale 15 ottobre 97 numero 26, articolo due oggetto, “promozione e valorizzazione della cultura e della lingua della Sardegna”.

<sup>27</sup> F. TOSO, *Il Patrimonio linguistico sardo* Parte VIII – Aspetti del panorama linguistico europeo, 20. LINGUE E DIALETTI– lingue e dialetti p. 5-6.

parlanti, o l'assenza di tradizioni letterarie, sono sufficienti a definire un idioma come dialetto<sup>28</sup>.

Oltre a essere strumenti di comunicazione, lingua e dialetto soddisfano anche un'altra importante funzione sociale: permettono il riconoscimento tra i membri di una comunità, i quali, attraverso l'adesione al medesimo codice comunicativo, stabiliscono un rapporto di affinità. Il fattore linguistico ha una funzione determinante nella definizione di un gruppo: l'idioma non rappresenta soltanto un aspetto della cultura di un popolo, ma è anche il veicolo per la trasmissione della cultura nel suo insieme. Il rispetto della diversità linguistica di una comunità (nazionale, regionale o locale) è un elemento fondamentale della valorizzazione del patrimonio culturale di cui tale comunità è portatrice<sup>29</sup>.

La definizione di un idioma come dialetto si collega inoltre all'esistenza sullo stesso territorio di un codice linguistico dominante dotato di maggior prestigio: la distinzione tra lingua e dialetto si pone quindi da un punto di vista politico-sociale. Il dialetto è espressione spontanea, non formalizzata, della cultura di una comunità, la lingua risponde invece alle esigenze di una società organizzata (e in particolare di uno stato) che al proprio bagaglio di consuetudini giuridiche, di storia comune e di tradizioni aggiunge un sistema di comunicazione istituzionalizzato fornito di una norma, accettato dai propri membri al di sopra delle (eventuali) varietà linguistiche. Ogni lingua parte dalle condizioni di dialetto e diventa lingua a tutti gli effetti quando ha il riconoscimento e il supporto di un potere politico<sup>30</sup>: «A language is a dialect with an army and a navy» recita un motto reso famoso dal sociolinguista Max Weinreich<sup>31</sup>. Resta difficile quindi affermare dove inizi una lingua e finisca un dialetto: non esiste una risposta facile, specialmente quando un idioma non è mai stata adottato da un governo moderno.

---

<sup>28</sup> M. ABLEY, op. cit., 2003, p. 48.

<sup>29</sup> F. TOSO, *Il Patrimonio linguistico sardo* Parte VIII – Aspetti del panorama linguistico europeo, 20. LINGUE E DIALETTI 6b.

<sup>30</sup> F. TOSO, *Il Patrimonio linguistico sardo* Parte VIII – Aspetti del panorama linguistico europeo, 20. LINGUE E DIALETTI p.7.

<sup>31</sup> M. ABLEY, op. cit., 2003, p. 140.



## II. 2. 2 L. N. 15 dicembre 1999 n. 482

Il contesto di applicazione della L. N. 26/1997 è stato modificato con l'entrata in vigore della L. N. 482/1999 che, in attuazione dell'articolo 6 della Costituzione, riconosce e tutela, tra le altre, la lingua e la cultura delle popolazioni catalane e di quelle parlanti sardo. Questo rappresenta un importante riconoscimento per la varietà sarda e per quella catalana e impone di approfondire le modalità di rapporto con le altre istituzioni interessate (Enti locali, Province, Istituti scolastici, Stato). Tale normativa, non contemplando alcune varietà linguistiche tutelate dalla normativa regionale (che la legge identifica come il tabarchino delle isole sulcitane e i dialetti sassarese e gallurese), apre una prospettiva di riflessione.

In merito a quanto detto è opportuno citare l'Art. 2, par.1 della Legge presa in esame:

In attuazione dell'articolo 6 della Costituzione e in armonia con i principi generali stabiliti dagli organismi europei e internazionali, la Repubblica tutela la lingua e la cultura delle popolazioni albanesi, catalane, germaniche, greche, slovene e croate e di quelle parlanti il francese, il franco-provenzale, il friulano, il ladino, l'occitano e il sardo<sup>32</sup>.

L'Art. 2 tutela un certo numero di varietà minoritarie tra cui il sardo sebbene escluda automaticamente il gallurese da misure specifiche di promozione e valorizzazione. Il Consiglio Provinciale di Olbia-Tempio attraverso la Delibera n. 44 del 19 settembre 2006 ha difeso la propria posizione delimitando l'ambito territoriale di tutela del sardo, specificando che nei comuni coinvolti la lingua parlata è il «Sardo nella variante Gallurese». Tale tentativo è insoddisfacente dal punto di vista formale e da quello di un'effettiva valorizzazione del patrimonio linguistico gallurese<sup>33</sup>.

Nonostante l'intento positivo della L. N. 482/1999 le sue scelte politiche compromissorie hanno rivelato forti limiti teorici e concettuali che si sono ripercossi anche sulle pratiche attuative. L'errore è stato quello di attribuire prerogative tipiche delle minoranze nazionali ad alcune minoranze linguistiche selezionate in base a criteri che tra hanno escluso altre realtà appartenenti a tale categoria, come la lingua zingara, il gallosiculo, il tabarchino e il sardocòrso, le ultime due contraddittoriamente riconosciute invece dalla già citata L. R. 26/1997. L'elencazione delle lingue tutelate di

---

<sup>32</sup> L. N. 482/1999, Art. 2.

<sup>33</sup> F. TOSO, *La Sardegna che non parla sardo*, Cagliari 2012, CUEC Editrice, p. 41.

fatto stabilisce una gerarchia all'interno del repertorio linguistico italiano; in altre parole, quanto del patrimonio linguistico nazionale non è incluso nella proposta all'art. 2 della Legge non risulta suscettibile di tutela poiché al di fuori di ogni categorizzazione. Attraverso la L. N. 482/1999 viene legittimata una valutazione discutibile del patrimonio linguistico italiano stabilendo una sorta di scala di valori tra ciò che merita una forma di tutela e ciò che non ne merita alcuna. Il numero chiuso delle minoranze linguistiche storiche, attraverso l'elencazione imprecisa e contraddittoria, ingloba realtà sociolinguistiche diverse tra loro, col risultato di fornire soluzioni di tutela insufficienti e poco efficaci<sup>34</sup>.

## II. 2. 3 La Limba Sarda Comuna

Un fenomeno da non trascurare è quello della diglossia della lingua sarda e delle varietà alloglotte in rapporto con l'italiano che ne ha di fatto messo in pericolo la sopravvivenza. I linguisti concordano nel sostenere che la situazione in Sardegna si avvia da un regime piuttosto stabile di diglossia degli ultimi due secoli a uno di alalia che prelude alla scomparsa delle varietà alloglotte o a quelle originarie e di identità storica. È stato necessario intervenire con le metodologie internazionali proprie della disciplina della rivitalizzazione e pianificazione linguistica per evitare la scomparsa della lingua sarda nell'arco di qualche generazione<sup>35</sup>. Per far fronte a tali necessità la Regione Sardegna ha deciso di cedere alla via compromissoria elaborando a tavolino la Limba Sarda Comuna con la Deliberazione n. 16/14 del 18 aprile 2006 *“Limba Sarda Comuna. Adozione delle norme di riferimento a carattere sperimentale per la lingua scritta in uscita dell'Amministrazione regionale”*.

### La Deliberazione dichiara che

sono state elaborate delle norme di riferimento, da utilizzarsi in via sperimentale per l'uso scritto in uscita di alcuni atti dell'Amministrazione regionale e per la traduzione di norme e documenti particolarmente importanti come lo Statuto sardo e la legge regionale n. 26 del 1997; che le norme linguistiche di riferimento a carattere sperimentale per la lingua scritta dell'Amministrazione regionale

---

<sup>34</sup> F. TOSO, *Patrimoni linguistici e lingue minoritarie: la prospettiva europea e quella italiana*, “Annali della Facoltà di Lingue e Letterature Straniere dell'Università di Sassari”, Vol. 5 (2005 pubbl. 2009), p. 115-124. p. 118.

<sup>35</sup> Relazione di accompagnamento al D.D.L. “Norme per la tutela, valorizzazione e promozione della lingua sarda e delle altre varietà linguistiche della Sardegna”, Regione Autonoma della Sardegna, 1997, pp-5-6.

consentono di avviare un processo graduale mirante all'elaborazione di una Limba Sarda Comuna, con le caratteristiche di una varietà linguistica naturale che costituisca un punto di mediazione tra le parlate più comuni e diffuse e aperta ad alcune integrazioni volte a valorizzare la istintività del sardo e ad assicurare un carattere di sovramunicipalità e la semplicità del codice linguistico; che la Limba Sarda Comuna intende rappresentare una "lingua bandiera", uno strumento per potenziare la nostra identità collettiva, nel rispetto della multiforme ricchezza delle varietà locali<sup>36</sup>.

### La Deliberazione propone inoltre

di adottare la Limba Sarda Comuna, come norme linguistiche di riferimento a carattere sperimentale per la lingua sarda scritta in uscita dell'Amministrazione regionale e per la traduzione di propri atti e documenti ufficiali, fermo restando, come previsto dall'articolo 8 della Legge 482 del 1999, "il valore legale esclusivo degli atti nel testo redatto in lingua italiana"<sup>37</sup>.

La Regione ha inteso elaborare una Limba Sarda Comuna con le caratteristiche di una varietà linguistica naturale per costituire un punto di mediazione tra le parlate più comuni e diffuse, aperta ad alcune integrazioni volte per valorizzare l'istintività del sardo e assicurare un carattere di sovramunicipalità e semplicità del codice linguistico. La Limba Sarda Comuna è stata intesa come «lingua bandiera»<sup>38</sup> e strumento per potenziare un'identità collettiva. Inoltre secondo la relazione di accompagnamento alla Legge nell'individuazione di una Limba Sarda Comuna, ci si riferisce paradossalmente a una lingua unica ma composta di tante varietà. Pertanto in teoria non vuole sostituirsi né imporsi sugli altri idiomi alloglotti della Sardegna dal punto di vista linguistico e giuridico.

L'intenzione è stata quella di individuare un punto di mediazione fra le varietà più comuni, ottenuto con il confronto delle differenze all'interno della lingua<sup>39</sup>. L'esame dei fenomeni soprattutto fonetici, studiati e censiti, che danno ai parlanti la sensazione di una grande differenziazione e frammentazione, ha privilegiato gli elementi di convergenza e sistematicità tra le varietà, gli elementi comuni a tutto l'insieme del

---

<sup>36</sup> Deliberazione n. 16/14 del 18.4.2006.

<sup>37</sup> Ibid.

<sup>38</sup> Relazione di accompagnamento al D.D.L. "Norme per la tutela, valorizzazione e promozione della lingua sarda e delle altre varietà linguistiche della Sardegna", Regione Autonoma della Sardegna, 1997, p-16.

<sup>39</sup> La discussione sul problema della standardizzazione è ancora vivace tra gli esperti della materia. La scarsa cognizione dei problemi generali della linguistica e della rivitalizzazione delle lingue impedisce a larghi settori della popolazione e della stessa classe dirigente di addentrarsi in una materia considerata terreno per specialisti. È ancora una volta la differenza percettiva, inconscia o meno, tra lingua e dialetto che ostacola la comprensione e la piena condivisione di questa tematica Relazione di accompagnamento al D.D.L. "Norme per la tutela, valorizzazione e promozione della lingua sarda e delle altre varietà linguistiche della Sardegna", Regione Autonoma della Sardegna, 1997, p. 17.

sardo<sup>40</sup>. La difficoltà di mettere in pratica un valido approccio di tutela del sardo non può però essere risolta manipolando i dati scientifici e costruendo una lingua naturale, occorre invece riconoscere la specificità anche delle altre varietà, nel caso in esame del gallurese<sup>41</sup>.

Le differenziazioni fin qui trattate conducono al tema del plurilinguismo quale condizione tipica e generale delle comunità linguistiche<sup>42</sup>: va sottolineato come la mancanza di una distinzione chiara tra esigenze di tutela dei diritti e dei patrimoni linguistici abbia generato confusione, al punto che le raccomandazioni comunitarie e i provvedimenti legislativi nazionali e regionali non rispondono alle singole esigenze. Guardando gli elenchi delle varietà linguistiche ammesse a tutela nei diversi paesi dell'UE all'atto della sottoscrizione della *Carta Europea delle Lingue Regionali o Minoritarie* il quadro è piuttosto confuso. Una parte dell'intellettualità sarda è in disaccordo con le disposizioni normative prese finora elencate. Il conseguimento dell'autonomia provinciale, ottenuta nel 2001 con l'istituzione della circoscrizione di Olbia-Tempio, è coinciso con l'accentuarsi delle iniziative finalizzate all'affermazione di una varietà standard di sardo come lingua ufficiale della Regione Autonoma, e con la diminuzione dell'attenzione per le minoranze interne<sup>43</sup>.

In tal senso meritano attenzione i documenti della Consulta Intercomunale Gallura (Istituto di Studi e Ricerca sulla Lingua e la Cultura Gallurese)<sup>44</sup>, un organismo di raccordo tra 22 comuni corsòfoni istituito nel 2001: esso

ha inteso rivendicare i diritti della lingua gallurese nella consapevolezza che non esiste una lingua sarda, ma che esistono più lingue parlate dai Sardi, aventi tutte diritto di essere salvaguardate, onde evitare il paradosso che un domani i Galluresi si trovino ad essere considerati stranieri in patria (Documento dell'8 dicembre 2001)<sup>45</sup>.

---

<sup>40</sup> Relazione di accompagnamento al D.D.L. "Norme per la tutela, valorizzazione e promozione della lingua sarda e delle altre varietà linguistiche della Sardegna", Regione Autonoma della Sardegna, 1997, p. 17.

<sup>41</sup> F. TOSO, op. cit. 2012, p. 24.

<sup>42</sup> F. TOSO, *Il Patrimonio linguistico sardo* Parte VIII – Aspetti del panorama linguistico europeo, 20. LINGUE E DIALETTI p.8.

<sup>43</sup> F. TOSO, op. cit., 2012, p. 40.

<sup>44</sup> <http://www.consultaintercomunalegallura.net/>

<sup>45</sup> Comunicato dell'Assemblea della Consulta, sede della consulta del gallurese, Arzachena, 2001.

Inoltre, in un documento del 7 febbraio 2003 sostiene anche

che l'impostazione della cosiddetta Lingua Sarda Unificata, frutto non di processi naturali, storici e culturali, ma di scelte politiche, praticamente mutuata e fondata, con qualche commistione, su una parlata sarda egemone [...] creerebbe motivo di divisioni, di forte attrito e giustificate resistenze tra le comunità delle altre aree linguistiche della Sardegna – e soprattutto in Gallura – con guasti irrimediabili ed ulteriori ritardi allo sviluppo e all'inserimento dell'Isola nella comunità internazionale<sup>46</sup>.

E ribadisce

l'assoluta contrarietà a progetti che, nei fatti, determinerebbero la morte del Gallurese innescando reazioni pericolose e difficilmente governabili e costringendo i Galluresi a cercare altre strade ed interlocutori per difendere il loro patrimonio linguistico e culturale<sup>47</sup>.

La Giunta regionale inoltre giudica come punto critico rilevante il fatto che lo Stato italiano non abbia ancora di fatto ratificato la Carta Europea. La mancata ratifica comporta, in sede giuridica e legislativa, ai fini della protezione di alcune minoranze linguistiche tra cui quella sarda, notevoli problemi e limitazioni di carattere normativo che influenzano notevolmente la redazione del disegno di legge. La ratifica della normativa europea favorirebbe enormemente la protezione della lingua di minoranza storica della Sardegna in tutti gli ambiti della visibilità e dell'uso sociale: scuola, università, mass media, pubblica amministrazione.

L'esperienza italiana dimostra che per quanto riguarda la tutela, la promozione e la valorizzazione del patrimonio linguistico tradizionale è opportuno riformulare il concetto di bene linguistico, eliminando la distinzione tra ciò che è meritevole di tutela e ciò che non lo è. Le forme di tutela dei patrimoni linguistici che si integrano in ciascun contesto nazionale andrebbero differenziate in base alle tipologie e alle situazioni, e in questo senso la tutela di ciascuna varietà minoritaria pone problemi specifici meritevoli di ulteriori approfondimenti<sup>48</sup>.

Anche l'elencazione delle lingue ammesse a tutela nelle leggi citate genera confusione e induce a equivoci, soprattutto perché molti dei provvedimenti previsti implicano già in partenza situazioni di standardizzazione linguistica e di accettazione di una varietà sovraordinata. Nel caso specifico sardo la frammentazione dialettale è uno

---

<sup>46</sup> Documento sulla lingua sarda unificata, Arzachena, 2003.

<sup>47</sup> Ibid.

<sup>48</sup> F. TOSO, (2009) *Patrimoni linguistici e lingue minoritarie: la prospettiva europea e quella italiana*. Annali della Facoltà di Lingue e Letterature Straniere dell'Università di Sassari, Vol. 5 (2005 pubbl. 2009), p. 115-124. ISSN 1828-5384. Articolo, p.122.

strumento identitario di notevole portata: in molti casi i parlanti tendono a riconoscersi non in un'identità linguistica astratta, sopralocale, bensì nello specifico della propria parlata materna. Da qui il rifiuto di proposte volte ad affermare il principio di una lingua comune a discapito di usi ancora vitali in ambiti più ristretti. Ai tentativi di affermazione di una norma comune si oppone quindi il problema del rispetto di una varietà interna che è spesso l'elemento che vincola effettivamente i parlanti all'utilizzo dell'idioma minoritario<sup>49</sup>.

La normativa, in generale, rivela una mancata consapevolezza della realtà e del vissuto storico delle comunità alloglotte e dei singoli individui che vi si integrano, accreditando le istanze di quei gruppi radicali dell'attivismo politico-culturale che interpretano il monolinguisimo a base minoritaria come meccanismo di acquisizione di una coscienza etnica del tutto slegata dal vissuto personale e dal retroterra culturale dei singoli parlanti<sup>50</sup>.

## II. 3 Il ruolo della linguistica computazionale

Come affermato nel precedente paragrafo, la legislazione attuale non fornisce i mezzi necessari per mantenere in vita la lingua sarda e in particolare la lingua gallurese, così come la sola pratica orale non è sufficiente ad accrescere la vitalità di una lingua<sup>51</sup>. In linea teorica, una lingua tradizionale forte in una comunità dovrebbe essere equipaggiata a resistere alla modernità o meglio ad avvantaggiarsene traendo guadagno dal beneficio dello sviluppo<sup>52</sup>. La tecnologia può dare un importante contributo in materia di tutela delle minoranze linguistiche per usufruire delle capacità di chi può normare una lingua in pericolo e preservarla per il futuro. I media sociali possono offrire per di più un'alta accessibilità sia per comunicazioni personali sia all'interno del gruppo minoritario. Le conoscenze e le competenze esistenti nel settore sono considerevoli e dovrebbero essere sfruttate; i finanziamenti europei per la ricerca e lo sviluppo hanno un impatto significativo in questo ambito. Le comunità che parlano una lingua a rischio di estinzione sarebbero così incentivate a promuovere il proprio idioma: risulta fondamentale un approccio mirato all'uso delle lingue in pericolo tra i giovani.

---

<sup>49</sup> F. TOSO, (2009) *Patrimoni linguistici e lingue minoritarie: la prospettiva europea e quella italiana*. Annali della Facoltà di Lingue e Letterature Straniere dell'Università di Sassari, Vol. 5 (2005 pubbl. 2009), p. 115-124. ISSN 1828-5384. Articolo, p.119 .

<sup>50</sup> Ivi, p.122.

<sup>51</sup> M. ABLEY, op. cit., 2003, p. 38.

<sup>52</sup> Ivi, p. 20.

La veicolare nei media rappresenta un'ancora di salvezza per le lingue minoritarie<sup>53</sup>: la presenza delle lingue minacciate in radio, TV, giornali o internet contribuisce naturalmente alla percezione che la lingua esiste, è viva e può partecipare alla vita quotidiana della popolazione. La presenza nei media potrebbe modernizzare il concetto di tutela: la comunicazione in lingua sarda sui media necessita di interventi mirati al fine di ottimizzare la destinazione delle risorse e raggiungere più facilmente l'obiettivo della visibilità e della qualità dei prodotti.

La letteratura scientifica e la sociolinguistica concordano nell'affermare che il messaggio da trasmettere è che le minoranze italiane ed europee rappresentano un interesse e una ricchezza per tutti. A questo proposito è bene ricordare che l'art. 22 della L. N. 482/1999 consente alla Regione Sardegna di incentivare e sostenere la presenza della lingua sarda nell'ambito delle tecnologie informatiche, in particolare su internet, in formato testuale e audiovisivo<sup>54</sup>.

L'UE in merito a tale strategia potrebbe giocare un ruolo centrale: nel suo intento di promuovere la condivisione delle buone prassi, dovrebbe incoraggiare tutti gli Stati membri a realizzare piani strategici nazionali per la promozione delle lingue a rischio di estinzione rifacendosi all'ottima qualità delle buone prassi già diffuse in alcune comunità linguistiche d'Europa. Il Consiglio d'Europa potrebbe poi valutare l'inserimento nella *Carta europea delle lingue regionali o minoritarie* delle clausole specifiche per le lingue in pericolo. Si potrebbero esortare gli Stati membri che ancora non l'hanno fatto a firmare la Carta, e gli Stati membri che l'hanno firmata a ratificarla. L'UE dovrebbe raccomandare agli Stati membri di offrire un sostegno simile a tutte le comunità che nel loro territorio parlano una lingua in pericolo o minoritaria. Nel definire le priorità applicabili a tutta Europa per la rivitalizzazione delle lingue, occorre porre l'accento sulla trasmissione delle lingue in pericolo in famiglia e sull'apprendimento delle stesse all'interno del sistema scolastico; sono necessarie quindi politiche solide nel campo dell'istruzione. I principali attori coinvolti, poi, dovrebbero collaborare al meglio nel sostenere le lingue a rischio di estinzione: le reti

---

<sup>53</sup> Relazione di accompagnamento al D.D.L. "Norme per la tutela, valorizzazione e promozione della lingua sarda e delle altre varietà linguistiche della Sardegna", Regione Autonoma della Sardegna, 1997 P. 24.

<sup>54</sup> L. N. 482/1999, "Interventi nel settore dei mezzi di comunicazione, internet e nuove tecnologie", Capo IV.

esistenti a livello europeo devono essere rafforzate e dotate di risorse sufficienti per operare in modo efficace ed efficiente in tale contesto<sup>55</sup>.

Nella Deliberazione n. 20/15 del 9 maggio 2005 in materia di *Promozione e valorizzazione della cultura e della lingua della Sardegna. Indagine socio-linguistica sulla lingua sarda* si dichiara:

confermando, dunque, la volontà di promuovere e tutelare, anche attraverso il sostegno della sperimentazione dell'insegnamento e dell'uso nelle scuole, a partire da quelle dell'infanzia e primarie, tutte le varietà linguistiche (sardo campidanese, sardo logudorese, sardo nuorese, catalano, tabarchino, sassarese, gallurese), occorre che l'Amministrazione regionale valuti, sulla base dei risultati di un'indagine scientifica, la possibilità di definire un codice linguistico in uscita rivolto a tutto il territorio regionale[...]<sup>56</sup>.

Altro compito della Commissione sarà di definire norme ortografiche comuni per tutte le varietà linguistiche in uso nel territorio regionale che aiutino a promuovere la creazione di word processor, correttori ortografici e l'utilizzo e la diffusione di strumenti elettronici per favorire l'uso corretto della lingua sarda<sup>57</sup>.

Si rivela quindi nodale in materia di tutela l'interazione tra la linguistica teorica (e applicata) e l'informatica, ossia tutti quegli strumenti tecnologici che si stanno rapidamente sviluppando nel campo della linguistica computazionale. È fondamentale il punto di incontro tra la ricerca linguistica e le tecnologie informatiche: l'analisi e l'esplorazione dei corpora, la lessicografia computazionale, il *Natural Language Processing (NLP)*, la traduzione automatica e la tecnologia del parlato<sup>58</sup>. L'informatica, ma anche la statistica, la matematica e l'intelligenza artificiale forniscono strumenti e metodi per le analisi linguistiche e per le loro applicazioni in campo computazionale, coprendo settori come la fonologia, la morfologia, la sintassi e la semantica<sup>59</sup>.

La regione Sardegna dovrebbe chiedersi in che modo il *computing* possa essere messo a servizio dell'indagine linguistica: il web e l'evoluzione telematica hanno reso molto più fluido il confine tra lingua scritta e lingua parlata. La posta elettronica, le chat, i blog e gli sms rappresentano alcuni dei nuovi stili di comunicazione in cui lo strumento elettronico e la forma dell'interazione fanno sì che il testo scritto contenga in realtà tratti lessicali, sintattici ed espressivi tipici della lingua parlata. I metodi e gli strumenti computazionali possono dunque essere applicati a tutta una serie di prodotti

---

<sup>55</sup> M.P. JONES, op. cit., 2013, pp.7-8.

<sup>56</sup> Deliberazione N. 20/15/2005, in materia di "Promozione e valorizzazione della cultura e della lingua della Sardegna. Indagine socio-linguistica sulla lingua sarda", p. 2.

<sup>57</sup> Ibid.

<sup>58</sup> I. CHIARI, *Introduzione alla linguistica computazionale*, Bari, Editori Laterza, 2007, Nota dell'autrice.

<sup>59</sup> Ivi, p.4.



del linguaggio rappresentativi delle modalità comunicative delle varietà linguistiche e dei registri delle varietà alloglotte sarde.

Alcuni Stati membri dell'UE hanno elaborato, soprattutto a livello di governi locali, piani strategici ben strutturati per la promozione e la salvaguardia delle loro lingue, in particolare le regioni spagnole e britanniche. In Irlanda, ad esempio, è stata sviluppata una strategia ventennale per la promozione dell'uso della lingua irlandese<sup>60</sup>. Negli ultimi decenni numerose innovazioni hanno promosso le lingue in pericolo e minoritarie, tra cui programmi condotti in comunità molto diverse come quella Sami<sup>61</sup>, quelle dell'Isola di Man<sup>62</sup> e del Galles<sup>64</sup>. Un progetto particolarmente riuscito è stato il MELT<sup>65</sup>, cofinanziato dall'UE, che ha promosso lo sviluppo di competenze per l'apprendimento prescolare delle lingue minoritarie<sup>66</sup>.

L'era digitale rappresenta una sfida per le lingue a rischio di estinzione, così come rappresenta anche un'opportunità. Una lingua minoritaria, in genere, è anche una lingua con poche risorse spendibili nell'informatica, sia a livello economico sia umano: la lingua araba ne è un esempio; nonostante questa sia una lingua largamente usata ha pochi strumenti e risorse pubblicamente disponibili. L'arabo manca di corpora, lessici, dizionari macchina e strumenti fondamentali di NLP completamente automatizzati come *tokenizers*<sup>67</sup>, *taggers*<sup>68</sup> e *parsers*<sup>69</sup>. In un recente articolo El-Haj et al. affermano l'esistenza di tre differenti paradigmi adatti alla creazione di risorse linguistiche<sup>70</sup>:

---

<sup>60</sup> <http://www.per.gov.ie/en/20-year-strategy-for-the-irish-language-2010-2030/>

<sup>61</sup> [https://www.regjeringen.no/globalassets/upload/fad/vedlegg/sami/hp\\_2009\\_samisk\\_sprak\\_engelsk.pdf](https://www.regjeringen.no/globalassets/upload/fad/vedlegg/sami/hp_2009_samisk_sprak_engelsk.pdf)

<sup>62</sup> <https://www.gov.im/>

<sup>63</sup> Per ulteriori approfondimenti cfr. M. ABBLEY, op. cit., 2013, p. 95.

<sup>64</sup> <http://gov.wales/topics/welshlanguage/promoting/capital-investment-fund-2015-to-2016/?lang=en>

<sup>65</sup> [http://info.melt-project.eu/ww/en/pub/melt\\_project/welcome.htm](http://info.melt-project.eu/ww/en/pub/melt_project/welcome.htm)

<sup>66</sup> M. P. JONES, op. cit., 2013, p.7.

<sup>67</sup> Il processo di segmentazione del testo in *token* è detto *tokenizzazione* e può essere realizzato automaticamente attraverso l'uso di appositi programmi chiamati *tokenizzatori/tokenizers*. In linguistica computazionale le unità di base del testo digitale sono i *token*, i quali raggruppano oltre alle parole ortografiche tradizionali anche numeri, sigle, segni di punteggiatura, nomi propri e altro. ALESSANDRO LENCINET AL., *Testo e Computer*, Carocci editore, Roma, 2007, p. 102.

<sup>68</sup> Il *part-of-speech (POS) tagging*, ossia l'etichettatura per categorie grammaticali, individua la parte del discorso che una parola ricopre in una data fase in input. Il *tagger* è il dispositivo computazionale che opera un *POS tagging* su materiale testuale. Questo riceve in input una frase e restituisce in output le forme grafiche delle parole accompagnate da etichette che segnalano la categoria grammaticale di appartenenza. I. CHIARI, op. cit., 2007, pp. 105-106.

<sup>69</sup> Un *parser* individua le relazioni tra i costituenti della frase, le loro gerarchie e attribuire a ogni forma della frase un'etichetta di categoria grammaticale. In generale, il *parsing* è un termine che indica un'analisi linguistica: esistono infatti *parsing* morfologici, sintattici e semantici. I. CHIARI, op. cit., 2007, pp. 101-102.

<sup>70</sup> M. EL-HAJ, U. KRUSCHWITZ, C FOX, *Creating language resources for under-resourced languages: methodologies, and experiments with Arabic - Language Resources and Evaluation*, 2015, 49.3: 549-580.

- il *crowdsourcing*<sup>71</sup>, per produrre una piccola ricerca rapidamente e in modo economico;

- la traduzione di uno standard di riferimento già esistente, processo facile ma potenzialmente di bassa qualità;

- l'utilizzo dei mezzi manuali con partecipanti umani competenti per creare una risorsa di alta qualità ma più dispendiosa.

Si rivela comunque possibile creare risorse supplementari in modo rapido ed economico anche con pochi mezzi; questo è vitale anche nel miglioramento degli strumenti NLP e delle applicazioni. La mancanza di risorse è invece un problema per quanto concerne la valutazione, cioè quando è importante disporre di metriche esterne in base alle quali valutare differenti metodi e tecnologie. Lo studio sulla lingua araba ha dimostrato che i tre diversi paradigmi sono utili per creare uno standard di riferimento per le lingue senza risorse, nello specifico la creazione di *corpora* per riassumere testi in lingua araba<sup>72</sup>.

In generale, le *Human Language Technologies (HLT)* sono interessate soprattutto a lingue che dispongono di grandi risorse o che sono di interesse per la scena politica o economica; al contrario, le lingue dei paesi in via di sviluppo o le minoranze linguistiche vengono meno trattate. Per colmare questa lacuna si potrebbe indirizzare la ricerca verso portabilità delle HLT per applicazioni multilinguistiche. Tale argomento è diventato essenziale negli ultimi anni. Il SALTIMIL (*Speech and Language Technology for Minority Languages*)<sup>73</sup>, un gruppo di interesse speciale dell'ISCA (*International Speech Communication Association*)<sup>74</sup>, è stato creato per promuovere la ricerca e lo sviluppo nel campo del parlato e della tecnologia linguistica per le lingue meno usate, in particolare quelle dell'Europa. Secondo il SALTIMIL le lingue minoritarie sono soprattutto «le lingue parlate da una minoranza di persone», ma nello studio l'attenzione verte su lingue che hanno «poche risorse spendibili nell'HLT». Queste lingue sono

---

<sup>71</sup> Cfr. infra CAPITOLO 3.

<sup>72</sup> M. EL-HAJ, U. KRUSCHWITZ, C FOX, *Creating language resources for under-resourced languages: methodologies, and experiments with Arabic - Language Resources and Evaluation*, 2015, 49.3: 549-580.

<sup>73</sup> <http://www.cstr.ed.ac.uk/~briony/SALTIMIL/>

<sup>74</sup> <http://www.isca-speech.org>

soprattutto quelle di paesi in via di sviluppo parlate invece da un ampio numero di persone: in particolare la lingua vietnamita conta circa 70 milioni di parlanti ma non ha risorse elettroniche<sup>75</sup>.

In aggiunta, l'*Automatic Speech Recognition* (ASR) include diversi aspetti<sup>76</sup>: 1. portabilità di modelli acustici; 2. modelli linguistici per nuove lingue; 3. costruzione di dizionari. Per costruire un modello linguistico e filtrare i documenti è necessario avere un vocabolario che contenga le parole da includere al modello. Ci sono molti metodi per costruire un dizionario, ad esempio si può usare un dizionario bilingue o multilingue. La costruzione di una base lessicale per una nuova lingua può avere un approccio collaborativo o si possono utilizzare dizionari già esistenti. Uno dei numerosi progetti in questo campo, denominato PAPHON, ambisce alla creazione di un *database*<sup>77</sup> lessicale multilingue che copra, tra le altre, la lingua inglese, francese, giapponese, malay, lao, thai e vietnamita. Dal progetto è stato elaborato un dizionario per la lingua vietnamita (francese-vietnamita e vietnamita-francese). Successivamente questo è stato filtrato fino a ottenere una lista di più di 40,000 entrate: collocazioni, prestiti e parole isolate<sup>78</sup>. Nello modello è stata fornita una metodologia efficace per la costruzione veloce di un modello per le lingue minoritarie: tale metodologia è stata testata e validata usando la lingua minoritaria vietnamita. In prima analisi sono stati costruiti strumenti per la lingua maggioritaria francese ed è stato validato il modello linguistico *web-based* a paragone con il classico modello linguistico *newspaper-based*. In seconda analisi è stato adattato lo strumento alla lingua vietnamita e sono stati definiti i moduli fissati e quali quelli specifici della lingua di arrivo. Raccogliendo regolarmente due siti di notizie quotidiane, in soli tre mesi è stato ottenuto un modello linguistico per la lingua vietnamita<sup>79</sup>.

Recenti lavori hanno stabilito ancora l'efficacia dell'Amazon's Mechanical Turk (MTurk)<sup>80</sup> per la costruzione di corpora paralleli per ricerche sulla traduzione automatica. Questo è stato applicato per la costruzione di una collezione di corpora

---

<sup>75</sup> V. B. LE ET AL. *Using the web for fast language model construction in minority languages* in "Interspeech", 2003.

<sup>76</sup> Ibid.

<sup>77</sup> Un database (base di dati) è una collezione di dati correlati creata con lo scopo specifico di rappresentare adeguatamente e memorizzare le informazioni relative a una realtà di interesse. Silvana CASTANO ET AL., *Informazione, conoscenza e Web per le scienze umanistiche*, Pearson Addison Wesley, 2009, p.2.

<sup>78</sup> V. B. LE ET AL. *Using the web for fast language model construction in minority languages* in "Interspeech", 2003.

<sup>79</sup> Ivi, p. 3-4.

<sup>80</sup> <https://www.mturk.com/mturk/welcome> (Cfr. infra CAPITOLO 3).

paralleli tra inglese e sei lingue dal subcontinente indiano: bengali, hindi, malayalam, tamil, telugu, e urdu<sup>81</sup>. Anche queste lingue hanno poche risorse, sono poco studiate, e mostrano fenomeni linguistici che sono difficili per le macchine di traduzione.

La qualità dei sistemi di Machine Translation (MT) statistici è fortemente correlata all'ammontare di testi paralleli disponibili per le coppie linguistiche. Comunque, la maggior parte delle lingue ha pochi dati di *training* bilingui immediatamente disponibili. Come risultato, la ricerca contemporanea MT tende a focalizzarsi sulle coppie linguistiche che già dispongono di una grande quantità di dati paralleli. La conseguenza è che le lingue che manifestano certi fenomeni linguistici particolari finiscono per non essere rappresentati, si tratta di lingue con morfologia complessa e lingue con ordini di parole divergenti. Matt Post et al.<sup>82</sup> descrivono il lavoro di raccolta e di perfezionamento di corpora paralleli a livello di documento tra l'inglese e ognuna delle sovra citate lingue *subject-object-verb* (SOV) del subcontinente indiano. Le lingue indiane sono importante oggetto di studio perché hanno scarse risorse in termini di disponibilità di sistemi MT (e strumenti NLP in generale) sebbene continuo mezzo miliardo di parlanti. Dato il crescente grado di penetrazione di internet in India, è importante trovare e impiegare traduttori inesperti attraverso servizi di crowdsourcing come MTurk. In ordine: 1. viene applicato un protocollo stabilito all'uso di MTurk per raccogliere dati paralleli per allenare e valutare sistemi di traduzione per sei lingue indiane; 2. vengono investigate le relative performance di modelli di traduzione sintattica su quelle gerarchiche; 3. si esplora l'impatto della qualità dei *training data* sulla qualità del modello risultante. I corpora paralleli sono stati raccolti usando un processo a tre passi disegnato per assicurare l'integrità della traduzione non professionale. Il primo passo è stato costruire un dizionario bilingue. I dizionari sono stati usati come mezzo di controllo sperimentale nella raccolta di quattro traduzioni di ogni frase della lingua di partenza. Al fine di misurare la qualità dei dati, sono stati dati dei voti per stabilire quale tra le quattro traduzioni ridondanti fosse la migliore. I dizionari sono stati poi usati per produrre glosse delle frasi della lingua di partenza, comparate poi alla traduzione prodotta dagli strumenti. È stata descritta la raccolta di sei corpora paralleli che contengono quattro traduzioni ridondanti del testo madre. Le

---

<sup>81</sup> M. POST, C. CALLISON-BURCH, M. OSBORNE, *Constructing parallel corpora for six indian languages via crowdsourcing* in "Proceedings of the Seventh Workshop on Statistical Machine Translation", Association for Computational Linguistics, 2012. p. 401-409.

<sup>82</sup> Ibid.

lingue indiane di questi corpora mostrano proprietà linguistiche marcatamente differenti se messe a paragone con l'inglese. Questo esperimento ha suggerito un nuovo approccio che potrebbe migliorare la qualità dei modelli costruiti con i *database* relazionali.

Altro esempio di unione tra linguistica e informatica è EUSLEM *lemmatiser/tagger* per il basco<sup>83</sup>. Un *lemmatiser/tagger* automatico è uno strumento base per le applicazioni come l'indicizzazione automatica, analisi sintattica e semantica o analisi di corpora testuali. Il suo compito è fornire il corretto lemma di una parola testuale e la sua categoria grammaticale. Caratteristiche<sup>84</sup>:

- è uno strumento *general-purpose*;
- gestisce lemmi standard e non-standard;
- gestisce e lemmatizza parole composte;
- è basato su analisi morfologiche.

Sebbene il *lemmatiser/tagger* per il basco sia uno strumento *general-purpose*, è stato applicato alla lessicografia e un *corpus manager* è diventato parte dello strumento per la creazione semiautomatica dei lessici. Il *lemmatiser/tagger* è stato concepito come uno strumento di base per altre applicazioni linguistiche. Il progetto è stato portato avanti da due entità: un gruppo della *Computer Science Faculty of The Basque Country University* e UZEI<sup>85</sup>, un'associazione che lavora sulla terminologia e lessicografia del basco. Il background di questo progetto è il *Systematic Compilation of Modern Basque (EEBS) Project*. Lo scopo è stato compilare e lemmatizzare semiautomaticamente un corpus di tre milioni di parole del ventesimo secolo dei testi baschi. Dal 1987 al 1992 UZEI ha creato un nuovo database che consiste in tre milioni di parole lemmatizzate, il quale viene rinnovato annualmente. Il *lemmatiser/tagger* ha dimostrato grande utilità per la lemmatizzazione di questo corpus, come anche per la lemmatizzazione del

---

<sup>83</sup> I. ALEGRIA ET AL. *Robustness and customisation in an analyser/lemmatiser for Basque*, in "Proceedings of Workshop on Customizing knowledge in NLP applications". Third International Conference on Language Resources and Evaluation. 2002.

<sup>84</sup> I. ADURIZ ET AL. *EUSLEM: A lemmatiser/tagger for Basque*. Proc. Of EURALEX'96, 1996, 17-26.

<sup>85</sup> <http://www.uzei.eus/>

*General Basque Dictionary Corpus*<sup>86</sup>. Il database lessicale viene continuamente aggiornato con il contributo di fonti differenti (dizionari moderni, proposte lessicali dall'Accademia della lingua basca, feedback dallo XUXEN *spelling corrector's users*<sup>87</sup>). Inoltre si stanno perfezionando le categorie e il sistema delle classi per andare incontro alle necessità che continuano a emergere con i processi di disambiguazione. È stato disambiguato manualmente un corpus di più di 20,000 parole allo scopo di estrarre le prime tavole per la disambiguazione statistica e per testare il risultato dei processi di disambiguazione.

MORFEUS ha implementato il modello sin qui descritto in modo da correggere anche le varianti dialettali e gli errori degli utenti. Questo per migliorare la robustezza dell'analizzatore morfologico e adattarlo a nuovi ambiti, per aumentarne la flessibilità, la copertura e la precisione<sup>88</sup>. Il design proposto potrebbe essere interessante anche per il trattamento di altre lingue. I risultati della ricerca sono quindi incoraggianti.

L'esplosione quantitativa e qualitativa dei testi in rete, la possibilità di raccogliarli non solo per forme concordanti di parole che vi occorrono, ma per contenuti concettuali e ontologici sta dando nuovo impulso alle procedure automatiche di analisi di testi in lingue storico-naturali e alla linguistica computazionale<sup>89</sup>. Lo studio del linguaggio mediante il calcolatore può fare luce sul rapporto tra facoltà del linguaggio e cognizione contribuendo a ricomporre in una nuova sintesi la dualità tra conoscenza e uso, competenza ed esecuzione linguistica, rappresentazione e computazione.

In conclusione è utile ribadire che la promozione dell'apprendimento linguistico e della diversità linguistica è uno degli obiettivi dei programmi dell'UE; tutte le lingue hanno i requisiti per ottenere il sostegno nell'ambito di programmi specifici: lingue ufficiali, lingue regionali, minoritarie e degli immigrati. Sarebbe però opportuno rivedere i parametri di definizione del concetto di lingua, sia a livello europeo sia a livello nazionale, e i conseguenti metodi di tutela. Il presente elaborato vuole appunto sottolineare il fondamentale apporto delle risorse linguistiche informatizzate per le lingue minoritarie. Accessibilità, portabilità e democraticità fanno dell'informatica il

---

<sup>86</sup> Questo corpus viene usato con il EEBS corpus dall'Accademia della Lingua per la compilazione del dizionario unificato (in collaborazione con UZEI).

<sup>87</sup> <http://xuxen.eus/>

<sup>88</sup> I. ALEGRIA ET AL., *Robustness and customisation in an analyser/lemmatiser for Basque*, in "Proceedings of Workshop on Customizing knowledge in NLP applications". Third International Conference on Language Resources and Evaluation. 2002.

<sup>89</sup> I. CHIARI, op. cit., 2007, Prefazione TULLIO DE MAURO.

nuovo baluardo in tema di tutela delle varietà nazionali italiane, dei dialetti europei e delle minoranze linguistiche e alloglotte nazionali e comunitarie.

## Capitolo III

### La costruzione di un dizionario-macchina gallurese con Nooj



### III. 1 La linguistica e la lessicografia computazionale

La linguistica computazionale è una disciplina che si colloca a cavallo tra scienze umane e scienze esatte e che si occupa dello «studio dei sistemi di elaborazione dedicati alla comprensione ed alla generazione del linguaggio»<sup>90</sup>. Le ricerche di linguistica computazionale sono spesso intrecciate con quelle di intelligenza artificiale ma fanno parte della linguistica computazionale anche le ricerche che usano tecniche informatiche non intelligenti, come quelle di stilistica computazionale e in generale quelle basate sull'elaborazione (anche con strumenti statistici) di corpora lessicali, allo scopo di realizzare dizionari o condurre vari tipi di sperimentazioni che simulano tramite computer attività in linguaggio naturale anche complesse, tra cui la traduzione automatica e assistita, l'analisi testuale automatica, il *parsing*, il riconoscimento e la generazione automatica dei testi<sup>91</sup>.

La linguistica computazionale può essere definita quindi come lo studio del linguaggio con l'ausilio del calcolatore ed è impegnata in tutti i settori della ricerca linguistica teorica, dalla sintassi alla pragmatica all'analisi del discorso, attraverso la costruzione di sistemi che realizzano teorie linguistiche<sup>92</sup>. L'analisi testuale automatica è una disciplina che si è sviluppata all'interno della linguistica computazionale nel momento in cui quest'ultima si è staccata dall'ambito delle ricerche di intelligenza artificiale e ha cominciato a utilizzare software dedicati, creati dall'informatica, per impostare e portare a termine specifici procedimenti analitici come la lettura automatica dei corpora e il recupero delle informazioni. L'informatica ha quindi fornito alla linguistica basi stabili per sviluppare simulazioni di operazioni cognitive complesse, ad esempio per analizzare corpora di grandi dimensioni. Di grande importanza in questo campo le capacità di memorizzazione dei supporti magnetici e ottici come gli hard disk di nuova generazione, i CD-Rom e i DVD-Rom scrivibili e riscrivibili, i quali permettono di registrare e riutilizzare grandi quantità di dati, agevolando classificazione, analisi e gestione automatica<sup>93</sup>.

---

<sup>90</sup> R. GRISHMAN, *Linguistica computazionale*, Milano, Tecniche Nuove, 1988.

<sup>91</sup> G. DE BUERIIS ET AL., *Lessici elettronici e descrizioni lessicali, sintattiche, morfologiche ed ortografiche*, Salerno, Plectica Editrice, 2005, p.38.

<sup>92</sup> Cfr. G. L. BECCARIA, (a cura di), *Dizionario di linguistica*, Bologna, 1984, p. 446.

<sup>93</sup> G. DE BUERIIS ET AL., op.cit., 2005, p.37.

### III. 2 I dizionari

La linguistica computazionale indirizzata alla produzione di applicazioni di utilità generale si è concentrata sullo studio del lessico (lessicologia) e sulla lessicografia, branca della linguistica che si occupa della preparazione ed elaborazione dei dizionari. Il campo lessicografico è stato uno dei primi a interessarsi alla possibilità di sfruttare le risorse dei mezzi informatici, la lessicografia computazionale come anche asserito da Patrick Hanks<sup>94</sup> nell'*Oxford Handbook of Computational Linguistics* ha due compiti:

- ristrutturare e sfruttare i dizionari tradizionali a fini computazionali;
- usare le tecniche computazionali per compilare nuovi dizionari<sup>95</sup>.

Attualmente tutti i principali dizionari vengono distribuiti con una versione elettronica su CD-ROM e alcuni di essi danno anche la possibilità di consultazione online. In molti casi la versione elettronica è identica nelle potenzialità e nel disegno alla versione cartacea del dizionario, ma tra gli sviluppi della lessicologia e della lessicografia computazionale esistono applicazioni per l'elaborazione di dizionari per nuovi impieghi: costruire un dizionario elettronico infatti non consiste semplicemente nel copiare il contenuto di un dizionario tradizionale cartaceo in una forma leggibile e consultabile con un computer. Vi sono diverse forme di dizionari elettronici, alcune più elementari e altre più complesse che includono numerose informazioni linguistiche per ogni lemma. Un dizionario elettronico avanzato può contenere (e permettere l'interrogazione su) un numero di informazioni linguistiche nettamente superiore a un normale dizionario cartaceo: oltre alla pronuncia e alla categoria grammaticale, può includere anche informazioni sulle frequenze d'uso (marche d'uso o dati quantitativi su corpora), informazioni lessicali, esempi tratti da corpora, informazioni morfo-sintattiche avanzate, informazioni semantiche e sulla presenza di possibili collocazioni. Per tale ragione il termine dizionario elettronico viene oggi usato in una molteplicità di sensi

---

<sup>94</sup> G. DE BUERIIS ET AL., op.cit. , 2005, p. 49.

<sup>95</sup> I. CHIARI, *Introduzione alla linguistica computazionale*, Editori Laterza, Bari, 2007, p. 84.

diversi, come semplici liste di parole o versioni online di grandi dizionari di tradizione o anche prodotti creati e progettati per il web<sup>96</sup>.

Si possono idealmente individuare due categorie principali di dizionari elettronici<sup>97</sup>:

- *human-oriented electronic dictionaries*: ne costituiscono l'esempio più comune i CD-ROM distribuiti insieme al proprio corrispondente cartaceo o i dizionari consultabili e scaricabili da internet. Spesso possono svolgere ulteriori funzioni oltre al consueto accesso a lemmi, accezioni e definizioni. Questo tipo di applicazioni sono anche dette di *Computer-Aided Traditional Lexicography*, ossia di lessicografia tradizionale assistita dal computer. Tra i dizionari informatizzati possiamo includere anche i dizionari basati su corpora, che offrono l'accesso diretto a risorse autentiche di materiale linguistico e possono essere progettati interamente su materiale estratto da corpora di riferimento e i dizionari, glossari e database terminologici autonomi o associati di traduzione;

- *machine-oriented electronic dictionaries*: vi è una stretta relazione tra repertorio lessicale e applicazioni computazionali che svolgono operazioni facendo ricorso a tale repertorio. Nelle applicazioni *dictionary-based* il termine dizionario viene usato per descrivere repertori contenenti liste di parole (sotto forma di lemma o di parola testuale) accompagnate da una serie di informazioni linguistiche aggiuntive che servono all'applicativo per svolgere determinate operazioni (disambiguare la categoria grammaticale, riconoscere una forma fonetica o produrla in modo appropriato rispetto a un input, e così via). Il dizionario non contiene definizioni esaustive dei sensi e degli usi di una parola, ma contiene invece informazioni di carattere linguistico relative a specifici ambiti di applicazione (fonetico, sintattico, semantico). Anche il lessico di frequenza è una sorta di dizionario-macchina che, inizialmente nato come strumento di pura investigazione linguistica, si è evoluto per entrare a far parte di molte delle applicazioni computazionali di Natural Language Processing e di traduzione automatica. Questi dizionari elettronici sono detti anche *lexical database* o di risorse lessicali proprio perché possono contenere informazioni linguistiche e metalinguistiche diverse da quelle che usualmente si trovano nei dizionari tradizionali.

---

<sup>96</sup> I. CHIARI, op. cit., 2007, pp. 85-86.

<sup>97</sup> Ibid.

Un dizionario-macchina è tecnicamente un repertorio di lemmi o forme flesse di una lingua associate a una serie di informazioni linguistiche i cui dati sono resi disponibili all'accesso di una o più applicazioni computazionali. Tali dizionari servono ad applicazioni quali, ad esempio, il riconoscimento e la sintesi vocale, la correzione ortografica, il *tagging* e il *parsing* sintattico o la traduzione automatica. Tipologie più recenti di dizionari-macchina contengono informazioni di tipo semantico come nel caso delle ontologie o delle reti semantiche. Pur essendo in qualche modo dipendente da altri applicativi, il dizionario-macchina può nascere in modo svincolato da essi e cioè può essere progettato per venire in un secondo tempo adottato da una moltitudine di differenti applicativi. Sinteticamente, i principali domini di applicazione dei dizionari-macchina esistenti sono i seguenti:<sup>98</sup>

- applicazioni di *Natural Language Processing*: dizionari-macchina di tipo morfo-sintattico che servono per la costruzione di applicazioni di *parsing* sintattico, *tagging* grammaticale, correzione ortografica e grammaticale, sillabazione di documenti;
- applicazioni delle tecnologie del parlato: dizionari-macchina dotati di informazioni sia fonetiche (acustiche e uditive) che morfo-sintattiche; sono associati ai sistemi di sintesi del parlato (TTS), di riconoscimento vocale e di dialogo uomo-macchina;
- ontologie e basi di conoscenza: con l'ausilio di dizionari elettronici a base semantica sono motori strutturati per l'estrazione di rappresentazioni astratte delle nostre conoscenze (*knowledge bases*) da testi di vario tipo. Lo sviluppo di ontologie, ossia di sistemi che rappresentano la struttura delle nostre conoscenze, a sua volta viene integrato in applicazioni di NLP come la correzione ortografica (per esempio per la disambiguazione delle omografie) o grammaticale, il riconoscimento del parlato, ma anche nella disambiguazione semantica, nel reperimento di informazioni nei documenti (*text mining*) e nell'estrazione di parole-chiave e riassunti da un testo (*summarization*);

---

<sup>98</sup> I. CHIARI, op. cit., 2007, p. 89.

- traduzione automatica: dizionari-macchina con informazioni lessicali e sintattiche (oltre che statistiche) indispensabili per i sistemi di traduzione assistita sia di tipo basato su regole che probabilistico;
- *information retrieval e information extraction*: i dizionari-macchina sono essenziali per una serie di motori di cui si fa uso comunemente per selezionare documenti che presentino determinate caratteristiche (come alcuni motori di ricerca sul web) e per estrarre specifiche informazioni rilevanti dagli stessi documenti o da database strutturati.

La struttura del materiale lessicale o sintattico che entra a far parte del dizionario deve essere resa esplicita e omogenea in modo da permettere un impiego uniforme, determinato e non ambiguo dell'applicativo cui si assocerà; un dizionario di questo tipo si differenzia rispetto a un dizionario tradizionale poiché contiene le forme cui l'applicazione dovrà accedere accompagnate da informazioni linguistiche aggiuntive in forma codificata.

La struttura e il tipo di informazioni che si possono trovare in un dizionario-macchina sono diverse, almeno in parte, da quelle che troviamo comunemente in un dizionario informatizzato<sup>99</sup>:

- un dizionario-macchina centrato sulla sintassi può contenere informazioni dettagliate sul comportamento di determinate forme linguistiche dal punto di vista morfo-sintattico, sugli usi idiomatici, sulle collocazioni, sulle più tipiche costruzioni sintattiche di cui entra a far parte ciascuna occorrenza di un lemma;

- un dizionario-macchina creato per uso fonetico o fonologico può contenere la specificazione di parametri per la sintesi o per il riconoscimento della forma da un applicativo di *text-to-speech* o per un sistema di dialogo uomo-macchina.

L'analisi testuale automatica del lessico-grammatica inaugurata dal linguista Maurice Gross per il francese e in seguito sviluppata e applicata all'italiano dal linguista Annibale Elia al Dipartimento di Scienze della Comunicazione dell'Università di

---

<sup>99</sup> I. CHIARI, op. cit., 2007, p. 86.

Salerno, è soprattutto di tipo morfosintattico, e si basa sull'uso di motori linguistici come i dizionari elettronici, i quali sono detti anche basi di dati lessicali. Attualmente esistono quattro pacchetti software che sfruttano questa modalità di strutturazione<sup>100</sup>:

- INTEX<sup>101</sup>: si basa sull'uso di dizionari elettronici, grammatiche locali e automi a stati finiti. La prima versione è stata realizzata agli inizi degli anni novanta da Max Silberstein sotto la guida di Maurice Gross. Il software è stato implementato fino al 2004, oggi è considerato non più implementabile;
- UNITEX<sup>102</sup>: è la versione speculare di INTEX, tuttora implementata per quanto riguarda il lingware;
- Nooj<sup>103</sup>: naturale implementazione di INTEX, del quale migliora alcuni aspetti relativi alla gestione della segmentazione delle parole in lettere o gruppi di lettere, nonché del rapporto tra parole semplici e parole composte;
- CATALOGA: realizzato da Alberto Postiglione (per la costruzione e la strutturazione della shell) e Mario Monteleone (per la gestione del lingware) del Dipartimento di Scienze della Comunicazione dell'Università di Salerno.

Le modalità di analisi di INTEX, UNITEX, e NOOJ sono molto simili tra loro e possono essere suddivise in tre fasi principali<sup>104</sup>:

- prima fase: il modulo iniziale effettua la lettura automatica di un testo, completandone sia l'indicizzazione sia la tokenizzazione, per poi eseguire il *matching* tra le parole contenute dal testo e le entrate catalogate e classificate nei dizionari elettronici. Il risultato del *matching* è la creazione di dizionari elettronici del testo analizzato. In questi dizionari le entrate sono elencate in ordine alfabetico e corredate da informazioni di tipo morfo-grammaticale e suddivise in base alla loro caratteristica di unità di significato autonome;

---

<sup>100</sup> G. DE BUERIIS ET AL., op. cit., 2005, p.39.

<sup>101</sup> Per ulteriori informazioni su INTEX, cfr. <http://intex.univ-fcomte.fr/>

<sup>102</sup> Per ulteriori informazioni su UNITEX, cfr. <http://igm.univ-mlv.fr/%7Eunitex/>

<sup>103</sup> Per ulteriori informazioni su NOOJ, cfr. <http://www.nooj4nlp.net/>

<sup>104</sup> G. DE BUERIIS ET AL., op. cit., 2005, p.40.

- seconda fase: l'analisi consente di leggere all'interno del testo effettuando specifiche ricerche visualizzabili sotto forma di concordanze e di effettuare la localizzazione di *pattern* sintattici, la disambiguazione e il *parsing* del testo. Queste operazioni sono possibili grazie all'uso di automi e trasduttori a stati finiti, realizzati sotto forma di grafi e applicati come elementi di lettura e analisi del testo. Tali automi vengono anche chiamati grammatiche locali, poiché in ognuno di essi vengono analizzate e descritte isolate caratteristiche morfosintattiche di una lingua, come ad esempio (per l'italiano) l'uso dei participi passati, l'accordo grammaticale o la forma passiva di specifici verbi. Al testo possono essere applicate contemporaneamente più grammatiche locali, o anche potenzialmente tutte le grammatiche locali realizzabili per la lingua italiana;
- terza fase: l'analisi consente di importare specifici file realizzati in forma di tabella con Microsoft Office Excel (soprattutto in INTEX e UNITEX) che contengono informazioni di carattere morfo-sintattico formalizzate in base alle proprietà distribuzionali e trasformazionali di ogni singolo elemento. Ogni singola tabella importata diventa una grammatica locale pronta per essere applicata durante l'analisi testuale automatica.

CATALOGA invece ha funzionalità e scopi diversi dagli altri tre software. Mutua da INTEX il *matching* fra testi e dizionari elettronici, ma si concentra sull'analisi terminologica del corpus, basata sulla localizzazione delle parole composte di tipo tecnico-scientifico. Questo software ha fasi di analisi diverse dai tre precedenti, riassumibili come segue<sup>105</sup>:

- lettura automatica dei testi;
- computo delle entrate composte terminologiche contenute nei testi;

---

<sup>105</sup> G. DE BUERIIS ET AL., op. cit., 2005, pp. 40-41.

- catalogazione automatica dei testi sulla base delle parole composte terminologiche da essi estratte, del loro numero, del settore della conoscenza più ricorrente, e tramite il computo statistico del rapporto rilevato tra parole terminologiche incluse;
- restituzione a video delle statistiche di occorrenza terminologica di ogni singolo testo, in ordine decrescente e suddivise in base al relativo settore della conoscenza.

Maurice Gross è stato anche uno dei linguisti che per primi hanno lavorato all'elaborazione dei modelli matematici con i quali sono state costruite le prime grammatiche formali idonee allo studio delle lingue naturali. Negli anni '70 cominciò a descrivere il comportamento dei verbi francesi che reggono una completiva e si accorse che il modello descrittivo di Rosenbaum<sup>106</sup>, costruito sull'inglese, non era sufficiente a tenere conto delle molteplici irregolarità che presentava il comportamento del francese. Gross esaminò in modo approfondito il comportamento di tutte le entrate lessicali verbali del francese e prendendo spunto da un'intuizione di Harris<sup>107</sup> elaborò uno schema matriciale per rendere conto di tutte le peculiarità del comportamento sintattico di ogni singolo verbo. Nacquero le tavole lessico-grammaticali, adattate al loro trattamento informatico, in cui in ogni rigo si trovava un verbo e in ogni colonna una proprietà combinatoria: al loro intersecarsi un + indicava l'accettabilità della proprietà, un - ne indicava la non accettabilità. Gross si concentrò sullo studio delle frasi semplici (elementari) inglobando in esse anche le frasi costruite intorno al verbo che reggeva una completiva considerata l'espansione di un sintagma nominale. Nel giro di alcuni anni di ricerca venne costituita una banca dati matriciale, che rendeva conto di più di duecento proprietà combinatorie relative a più di 10.000 usi verbali. Uno dei primi risultati di questa ricerca tassonomica fu la scoperta che non c'erano, in sostanza, due verbi che avessero lo stesso comportamento sintattico, anche se essi potevano essere classificati in insiemi più larghi con un comportamento omogeneo. Secondo lo studioso per la

---

<sup>106</sup> P. S. ROSENBAUM, *The Grammar of English Complement Constructions*, MIT Press, Cambridge, 1967, Mass.

<sup>107</sup> Z. S. HARRIS, *From Morpheme to Utterance*, in "Language", n. 22, 1946, pp. 161-183.



descrizione dei tipi di frasi e di sintagmi ad alto livello di fissità il modello ad automi finiti è estremamente utile e potente<sup>108</sup>.

Gross<sup>109</sup> ipotizzò anche la realizzazione di un dispositivo che potesse integrare il formalismo a stati finiti; tale dispositivo, che è equivalente al componente trasformazionale, agirebbe sulla rappresentazione a stati finiti ponendo delle restrizioni sulle strutture di frasi. Così costruì un progetto di descrizione formale del lessico e della grammatica del francese come modello per altre lingue e basato sulle seguenti componenti<sup>110</sup>:

- dizionario elettronico delle forme canoniche;
- dizionario elettronico delle forme flesse;
- dizionario elettronico delle forme composte canoniche;
- dizionario elettronico delle forme composte flesse;
- tavole lessico-grammaticali (verbi, aggettivi, nomi, determinanti, avverbi, preposizioni, congiunzioni);
- grammatiche locali (automi a stati finiti).

L'idea di utilizzare gli automi per descrivere una lingua naturale era già stata indicata da Chomsky<sup>111</sup>. In effetti, il linguista accennò al modello degli automi finiti essenzialmente per eliminarlo a favore dei livelli superiori della sua gerarchia. L'argomentazione di Chomsky per escludere gli automi finiti come modello adeguato delle lingue naturali è fondata sulla presenza di strutture incassate, come le preposizioni condizionali:

---

<sup>108</sup> S. VIETRI, *Dizionari elettronici e grammatiche a stati finiti*, Salerno, Plectica editrice, 2008, p.7.

<sup>109</sup> M. GROSS, *Local Grammars and their Representation by Finite Automata*, in Hoey M. (ed.), 1993, *Data, Description, Discourse. Papers on the English language in honour of John McH Sinclair*, Harpers Collins: London, pp. 26-38.

<sup>110</sup> S. VIETRI, op. cit., 2008, pp. 9-10.

<sup>111</sup> N. CHOMSKY, *The Logical Structure of Linguistic Theory*, 1975.

*se S1 allora S2*

Chomsky scartò tale modello perché allo stesso tempo troppo potente e troppo limitato per descrivere il funzionamento di una lingua naturale. Gross, invece, vide nel modello a stati finiti un ottimo formalismo per descrivere quei segmenti della lingua naturale che si presentano, appunto, limitati nella loro libertà combinatoria<sup>112</sup>.

In base ai dati fin qui esposti sui software di analisi testuale automatica risulta evidente che nessuno di essi potrebbe funzionare correttamente senza l'ausilio di dizionari elettronici. Per fornire una definizione chiara ed esaustiva di cosa siano esattamente i dizionari elettronici del lessico-grammatica e quali siano le loro caratteristiche strutturali è essenziale chiarire il concetto di base di dati, cioè chiedersi cosa sia una base di dati e quando un file da noi elaborato può essere definito tale<sup>113</sup>.

L'informatica stabilisce che una base di dati è un

insieme di informazioni esaurienti e non ridondanti necessarie a una serie di applicazioni automatizzate e conosciute da un insieme logico che ne garantisce la gestione [...] La formulazione di una base di dati [...] è un processo che partendo dall'osservazione di una situazione reale giunge alla definizione della base di dati corrispondente<sup>114</sup>.

Una base di dati è inoltre un

insieme, anche molto esteso, di informazioni di diverso tipo che fanno riferimento ad uno specifico settore di conoscenza o ad una determinata organizzazione. Questi dati sono organizzati secondo precisi criteri e mediante precise strutture informative, per poter essere consultati, aggiornati ed eventualmente ristrutturati per mezzo di procedure organizzate in modo unitario<sup>115</sup>.

In termini algebrici le basi di dati sono degli insiemi finiti poiché includono elementi con caratteristiche comuni tali da poter essere descritti congiuntamente secondo un unico metodo. La realizzazione di una base di dati è governata da rigorosi criteri di organizzazione formale che impongono la catalogazione dei contenuti secondo campi ed etichette univoci e non ambigui, da applicare a tutti gli elementi dell'insieme da descrivere. È questo tipo di impostazione che consente il trattamento delle basi di dati tramite computer, con la creazione e l'applicazione di *Data Base Management*

---

<sup>112</sup> S. VIETRI, op. cit., 2008, p.9.

<sup>113</sup> G. DE BUERIIS ET AL., op. cit., 2005, pp. 44-45.

<sup>114</sup> P. MORVAN, *Dizionario di informatica*, Roma, Gremese-Larousse, 1989.

<sup>115</sup> N. BARCELLONA, M. MARINI, MONTI, P.,VERCESI, *5000 termini dell'informatica*, Gruppo Editoriale Jackson, Milano, 1988, p. 106.

*System* (DBMS) mirati che permettono di effettuare rapidamente operazioni di consultazione, *information retrieval*, importazione ed esportazione verso altre basi di dati, aggiornamento e copia<sup>116</sup>.

I dizionari elettronici vengono realizzati in diversi modi<sup>117</sup>:

- in riferimento a un settore specifico della conoscenza, nel nostro caso il linguaggio naturale, e più specificamente il lessico di una lingua;
- con una strutturazione empirica basata su specifiche osservazioni di aspetti e fenomeni reali, ovvero quelli relativi all'uso del lessico;
- con criteri di strutturazione omogenei, in quanti gli elementi che include sono le unità lessicali di una stessa lingua e hanno quindi per definizione delle caratteristiche in comune;
- in funzione di un uso esclusivamente informatico e con un rigoroso formalismo descrittivo che ne renda lo stoccaggio su supporti magnetici al fine di essere impiegato solo in sistemi logici di gestione, ovvero da software di analisi linguistica.

I dizionari elettronici dell'italiano, elaborati su base lessico-grammaticale, sono essenzialmente di due tipi, suddivisibili in base all'aspetto formale e semantico del loro contenuto<sup>118</sup>:

- dizionari elettronici di parole semplici (denominati DELAS-DELAF) che includono tutte le parole semplici dell'italiano, sia monotematiche che polirematiche;
- dizionari elettronici di parole composte (denominati DELAC-DELACF), qui anche definite parole composte polirematiche, ovvero come già indicato tutte

---

<sup>116</sup> G. DE BUERIIS ET AL., op. cit., 2005, p. 45.

<sup>117</sup> Ibid.

<sup>118</sup> G. DE BUERIIS ET AL., op. cit., 2005, p. 46.

quelle sequenze formate da due o più parole e che costruiscono congiuntamente singole unità di significato.

Tale suddivisione è necessaria sia dal punto di vista formale e morfologico, sia da quello semantico. Infatti, in fase di compilazione di un software query, recupero informazioni o analisi testuale automatica, come ad esempio i già citati INTEX, UNITEX, CATALOGA, l'assenza di separatori all'interno delle parole semplici e la presenza di questi all'interno delle parole composte, rappresentano fattori discriminanti e comportano impostazioni e scelte differenti per quanto riguarda gli automatismi nel trattamento dei dati. Nel caso delle parole semplici è infatti necessario prevedere e trattare solo dati alfabetici o numerici; con le parole composte, la presenza di separatori inserisce un livello aggiuntivo di dati, ai quali si dovranno assegnare funzioni univoche, non ambigue, e di valore diverso da quelli alfabetici e numerici<sup>119</sup>.

Le prospettive ipotizzabili per un simile approccio analitico e per un uso ampio di composti e polirematiche potrebbe essere<sup>120</sup>:

- l'inserimento di simili basi di dati lessicali all'interno di sistemi di interrogazione automatica, che possano essere utilizzati per scopi informativi diversi, come ad esempio quelli previsti dall'*e-government* e dalla cittadinanza elettronica;
- l'applicazione e l'uso dei concetti di unità lessicale e semantica delle polirematiche all'interno dei vari settori di ricerca legati alla strutturazione del web semantico, in particolare quelli che prevedono la creazione di ontologie non ambigue, quindi univocamente legate al materiale lessicale che definiscono. In tal senso, sempre nell'ambito della costruzione di ontologie possibili, i campi semantici delle polirematiche possono essere facilmente interpretati come delle macroetichette in cui inserire specificazioni più dettagliate.

---

<sup>119</sup> G. DE BUERIIS ET AL., op. cit., 2005, p. 46.

<sup>120</sup> Ivi, pp. 45-46.

### III. 1. 2 I dizionari nel Cloud<sup>121</sup>

Negli ultimi anni abbiamo assistito a un cambiamento radicale nella disponibilità di risorse linguistiche su internet dovuto al nuovo sviluppo della tecnologia Cloud. Nello specifico il Cloud comprende una serie di tecnologie che permettono di elaborare, archiviare e memorizzare una grande quantità di dati grazie all'utilizzo di risorse hardware e software disponibili in rete. Si utilizza un servizio cloud quando, tramite pc, smartphone o altri dispositivi vengono sfruttate delle risorse o dei servizi attraverso la rete. Le due tendenze attuali nello sviluppo e nell'offerta di prodotti e servizi su internet sono rappresentate dal *cloudcomputing* e dal *crowdsourcing*.

**Cloudcomputing.** Il termine cloudcomputing apparve negli Stati Uniti nel 2007 per designare l'abilità di archiviare e processare dati in uno spazio virtuale cui si può accedere in qualsiasi momento e in contemporanea con altri utenti. È una tecnologia basata su una rete di computer tra essi comunicanti e abilita gli utenti a un accesso ubiquo a piattaforme software, infrastrutture e servizi online. Generalmente la piattaforma viene offerta da grandi società quali, ad esempio, Google, Amazon e Microsoft Azure. Le soluzioni *cloud-based* hanno diversi vantaggi: non sono legate a specifici sistemi operativi, sono costantemente aggiornate senza costi aggiuntivi per gli utenti e sono accessibili ovunque e in qualsiasi momento su internet tramite PC e smartphones. Inoltre, l'uso del cloudcomputing per lo sviluppo, la manutenzione e la condivisione di dizionari e risorse terminologiche sta cambiando considerevolmente il panorama linguistico grazie alla costruzione di piattaforme online, all'interattività, all'interoperabilità tra gli utenti e alla dinamicità dei sistemi. La disponibilità di una nuova generazione di dizionari ha perciò incoraggiato l'adozione di nuove forme di cooperazione tra diversi attori nel processo: lessicografi, terminologisti, traduttori ed editori, ma anche utenti comuni, i quali agiscono attivamente tramite le risorse e gli strumenti disponibili in un ambiente condiviso, accorciando così il tempo di completamento di un progetto.

---

<sup>121</sup> J. MONTI, *Le tecnologie per la traduzione nell'era del cloud computing: stato dell'arte e prospettive future*, "Atti del III workshop internazionale Traduzione per le aziende, il territorio, l'editoria: teoria, prassi, didattica", Centro Linguistico di Ateneo Università di Salerno, 18-20 dicembre 2013, pp.95-109.

**Crowdsourcing.** Il crowdsourcing si riferisce al contributo di lessicografi professionisti e utenti comuni al fine di sviluppare risorse linguistiche sia dietro compenso sia su base volontaria. Il coinvolgimento degli utenti su vasta scala sta diventando una pratica comune nel processo di compilazione di risorse linguistiche. In questo modo l'utente passa da un ruolo passivo a uno attivo diventando un *prosumer*, termine che rappresenta la fusione tra *producer* e *consumer*<sup>122</sup>.

Grazie al crowdsourcing è possibile portare a termine grandi progetti di compilazione di dizionari in breve tempo e aggiornare costantemente le risorse linguistiche. Produrre tali risorse in modo manuale costituirebbe inoltre un'operazione dispendiosa, con il crowdsourcing, al contrario, è possibile una produzione di grandi volumi di dati in poco tempo, a costi contenuti, e con una discreta qualità. Questo nuovo approccio alle fasi di sviluppo, manutenzione e aggiornamento delle risorse lessicali e terminologiche, può essere identificato in:

- risorse lessicali per dizionari online, ossia piattaforme Wiki come ad esempio il Wiktionary<sup>123</sup>; l'UrbanDictionary<sup>124</sup>, un dizionario online uso contemporaneo di slang inglese; Lingro<sup>125</sup>, un dizionario collaborativo che combina dizionari aperti sul web con il contributo dell'utente; case editrici che pubblicano dizionari tradizionali adottano forme di collaborazione, Collins<sup>126</sup>, Oxford<sup>127</sup>, e Macmillan<sup>128</sup>;
- risorse terminologiche per database terminologici online, come TermWiki<sup>129</sup>, la controparte terminologica del Wiktionary<sup>130</sup>. TermWiki è basato su termini di uno specifico argomento ed è composto di brevi entrate di tipo enciclopedico. Le entrate vengono create e gestite dagli utenti che hanno così accesso istantaneo a termini in continuo aggiornamento; viene inoltre promossa la partecipazione

---

<sup>122</sup> R. LEW, *User-generated content (UGC) in English online dictionaries*, A. Abel & A. Klosa (eds.) *Ihr Beitrag bitte! – Der Nutzerbeitrag im Wörterbuchprozess (OPAL – Online publizierte Arbeiten zur Linguistik)*, Institut für Deutsche Sprache, Mannheim, 2013.

<sup>123</sup> [https://en.wiktionary.org/wiki/Wiktionary:Main\\_Page](https://en.wiktionary.org/wiki/Wiktionary:Main_Page)

<sup>124</sup> <http://www.urbandictionary.com/>

<sup>125</sup> <http://lingro.com/>

<sup>126</sup> [http://www.collinsdictionary.com/cas/login?service=http%3A%2F%2Fwww.collinsdictionary.com%2Fspring\\_cas\\_security\\_check](http://www.collinsdictionary.com/cas/login?service=http%3A%2F%2Fwww.collinsdictionary.com%2Fspring_cas_security_check)

<sup>127</sup> <http://public.oed.com/appeals/>

<sup>128</sup> <http://mcmillandictionary.com/open-dictionary/latestEntries.htm>

<sup>129</sup> <http://it.termwiki.com/>

<sup>130</sup> <https://en.wiktionary.org/wiki/Wiktionary>

dell'utente, così come l'acquisizione e la condivisione di dati terminologici multilinguistici;

- risorse lessicali e semantiche per il Natural Language Processing (NLP) e per la traduzione automatica e assistita. Tramite specifiche piattaforme, come ad esempio Mechanical Turk<sup>131</sup>, vengono distribuiti compiti e progetti di largo sviluppo per lessicografi professionisti o occasionali.

**User.** Lo *user*, o utente, può apportare diversi tipi di contributo al contenuto lessicale per i dizionari online e i database terminologici o i tesauri<sup>132</sup>. Tali contributi sono stati analizzati da diversi studiosi: secondo FUERTES-OLIVERA<sup>133</sup>, ad esempio, vi è differenza tra *collective free multiple-language internet reference works*, come Wikipedia e Wiktionary, che sono interamente compilati dagli utenti senza controllo editoriale da parte di *institutional internet*, e *reference works* che sono invece offerti da editori ufficiali. LEW<sup>134</sup> opera la stessa distinzione introducendo però il concetto di *collaborative-institutional dictionaries*, offerto da case editrici che permettono i contributi diretti dell'utente. In seguito LEW<sup>135</sup> ha anche identificato diversi gradi di intervento da parte dell'utente nello sviluppo del contenuto lessicale facendo delle distinzioni: *collaborative dictionaries*, contenuti esclusivamente generati dagli utenti; combinazione tra contenuti generati dagli utenti (*user-generated*) e contenuti professionali; scambio con altri utenti di dizionari tramite forum online. Sulla base della differenziazione di LEW, ABEL e MEYER<sup>136</sup> hanno poi elaborato una successiva classificazione: *direct user contribution*, la quale include articoli interamente o parzialmente scritti da utenti in collaborazione; *indirect user contribution* che si manifesta in diverse forme di feedback esplicito (ad es. via email o altre forme nel web)

---

<sup>131</sup> <https://www.mturk.com/mturk/welcome>

<sup>132</sup> J. MONTI, op.cit., 2013, p. 98.

<sup>133</sup> P. A. FUERTES OLIVERA, *The Function Theory of Lexicography and Electronic Dictionaries: WIKTIONARY as a Prototype of Collective Free Multiple-Language Internet Lexicography*, H. Bergenholtz, S. Nielsen & S. Tarp (eds.) *Lexicography at a Crossroads: Dictionaries and Encyclopedias Today, Lexicographical Tools Tomorrow*, Bern, Peter Lang, 2009, pp. 99-134.

<sup>134</sup> R. LEW, *Online dictionaries of English*, P.A. Fuertes-Olivera & H. Bergenholtz (eds.) *e-Lexicography: The Internet, Digital Initiatives and Lexicography*, London/New York, Continuum, 2011, p. 230-250.

<sup>135</sup> R. LEW, op. cit. 2013.

<sup>136</sup> A. ABEL, C. M. MEYER, *The dynamics outside the paper: user contribution to online dictionaries*, Proceedings of the 3rd Biennial Conference on Electronic Lexicography (eLex), Ljubljana, Trojina, Institute for Applied Slovene Studies/Tallinn, Eesti Keele Instituut, October 2013, Tallinn, Estonia, pp. 179-194.

o feedback implicito (attraverso analisi file log o contenuti esterni *user-generated*); *accessory user contribution* che va oltre il contenuto del dizionario permettendo uno scambio sia tra l'editore del dizionario e gli utenti, sia tra gli stessi utenti.

Il contributo dell'utente dipende ancora dal tipo di dizionario e dal suo grado di apertura a forme collaborative:

1. Il contributo dell'utente nelle piattaforme lessicali/terminologiche Wiki-type come Wiktionary, Termwiki o UrbanDictionary è **diretto**, e avviene generalmente su base volontaria. I lessicografi/terminologisti volontari, professionisti e non, lavorano gratuitamente ricompensati da forme non-convenzionali di gratificazione sociale: la pubblicazione del loro nome nella lista dei contributori, il riconoscimento pubblico della loro leadership nel momento in cui raggiungono la cima della classifica, o semplicemente l'opportunità di acquisire nuove conoscenze e competenze. Nella contribuzione diretta l'utente ha la possibilità di editare direttamente l'entrata lessicale/terminologica sulla base del modello predefinito senza nessuna particolare supervisione da parte di editori professionisti<sup>137</sup>.

Wiktionary, ideato come partner lessicale di Wikipedia, è un esempio indicativo di dizionario edito in modo collaborativo: è un progetto Wikimedia Foundation<sup>138</sup> atto a creare un dizionario dal contenuto gratuito e multilingue. Il progetto è iniziato nel Dicembre 2002 ed è ora una dei più estesi dizionari nel cloud. Wiktionary è un wiki, ciò significa che gli utenti possono editarlo e che tutti i contenuti sono autorizzati sia dal Creative Commons Attribution-ShareAlike 3.0 Unported License<sup>139</sup> sia dal GNU Free Documentation License<sup>140</sup>. Attualmente Wiktionary è disponibile in più di 40 lingue che forniscono circa 14.1 milioni di articoli e vi si trovano numerose informazioni lessicali standard (classe lessicale, pronuncia, forme non flesse, etimologia, definizione del significato, proposizioni esemplificative e traduzioni), ma sono inclusi anche tesauri, poesie, frasari, traduzioni, file audio di pronuncia, etimologie, citazioni, statistiche linguistiche e appendici estensive. A titolo esemplificativo si può riportare l'Index:

---

<sup>137</sup> J. MONTI, op.cit., 2013 p.99 e segg.

<sup>138</sup> [https://wikimediafoundation.org/wiki/Pagina\\_Principale](https://wikimediafoundation.org/wiki/Pagina_Principale)

<sup>139</sup> <https://creativecommons.org/licenses/by-sa/3.0/it/>

<sup>140</sup> <http://www.gnu.org/gnu/thegnuproject.it.html>



Welsh<sup>141</sup>, una lista che contiene tutti i termini correttamente ordinati della lingua gallese.

La natura illimitata di Wiktionary comporta problemi principalmente correlati all'affidabilità e alla completezza delle informazioni lessicografiche: la mancanza di revisione editoriale non fornisce protezione contro informazioni obsolete, incorrette, incomplete, inaccurate o troppo complesse. Inoltre, il contenuto del dizionario è esposto a vandalismo e spam di internet. Anche il controllo della qualità viene affidato agli utenti che tramite feedback anonimi chiedono la verifica delle entrate lessicali (con la richiesta di citazioni) o la cancellazione delle stesse plausibili ragioni.

Numerosi dizionari wiki e basi terminologiche prevedono un apporto da parte dello *user*, ma in alcuni casi quest'ultimo può collaborare alla creazione dell'informazione lessicale o terminologica attraverso una modalità più strutturata rispetto al modello Wiktionary, è questo il caso di Termwiki. Questa è una rete considerevole di *social learning* che permette agli utenti di imparare, scoprire, condividere e raccogliere termini e glossari personali in 1487 settori e in 97 lingue. Gli utenti possono aggiungere termini tramite un input strutturato e con liste di scelta disponibili per molteplici categorie di dati, in modo da assicurarne la coerenza. In entrambi i casi, Wiktionary e Termwiki, i contributi dell'utente seguono un modello molto semplice in cui gli utenti prima creano le entrate del dizionario e successivamente le entrate possono essere editate dal contribuente iniziale o da altri utenti per varie ragioni.

Oltre ciò, l'innovativa piattaforma *cloud-based TaaS (Terminology as a Service)*<sup>142</sup> fornisce ai linguisti e alle applicazioni linguistiche servizi utili a favorire la creazione, la validazione, l'armonizzazione, la condivisione e l'applicazione delle risorse terminologiche. La sua caratteristica distintiva è la forte interazione tra utente umano e mezzi computazionali. Nel processo di sviluppo i termini candidati sono estratti automaticamente sia da raccolte di dati paralleli che comparati, estratti da corpora di giornali comparabili, Wikipedia multilingue, e da corpora comparabili focalizzati. In seguito la lista di coppie dei termini candidati viene estratta e immessa nel deposito termologico TaaS. Durante l'esecuzione del progetto terminologico i dati vengono proposti all'utente per la loro validazione. Tramite questa procedura il database

---

<sup>141</sup> <https://en.wiktionary.org/wiki/Index:Welsh>

<sup>142</sup> <https://demo.taas-project.eu/>

TaaS cresce costantemente e rapidamente attraverso la validazione dei dati raccolti dal web da parte dell'utente.

Altra interessante sperimentazione nell'uso delle tecniche di crowdsourcing è la piattaforma Language Depot<sup>143</sup>: questa viene fornita come servizio per le comunità linguistiche dal Language Software Group dell'Istituto Linguistico della Payap University in Thailandia. La piattaforma è usata principalmente dai parlanti di lingue con poche risorse interessati allo sviluppo di un dizionario e alla trascrizione/raccolta di racconti nella loro lingua. Tali lingue sono principalmente usate da comunità linguistiche ristrette e sono spesso impossibili da decifrare e certificare senza l'aiuto dei parlanti nativi. Gli utenti possono aggiungere la descrizione linguistica usando il programma open source WeSay<sup>144</sup>, sviluppato appunto per aiutare le comunità a compiere autonomamente attività di sviluppo dei loro idiomi. Il programma permette una facilitazione nella documentazione di parole ed espressioni attraverso un'interfaccia *easy-to-use* e prevede una divisione dei compiti tra i vari membri della comunità secondo le loro specializzazioni.

2. Le principali case editrici inglesi hanno introdotto risorse linguistiche crowdsourcing aprendo i loro dizionari a suggerimenti pubblici. In dizionari istituzionali, come il Collins o il Macmillan, gli utenti, generalmente non professionisti, contribuiscono **indirettamente** suggerendo nuove entrate. Il loro lavoro viene poi sottoposto a un processo di revisione e dopo una verifica da parte degli editori vengono caricati o meno, questo a seconda del tipo di dizionario.

Un caso è rappresentato dall'Oxford English Dictionary (OED) che nel 1859 ha dato la possibilità al pubblico di lingua inglese di fornire un contributo con esempi di parole in uso: in circa 70 anni, l'OED ha ricevuto 6 milioni di contributi<sup>145</sup>. I lettori dell'Oxford English Dictionary ora possono inviare informazioni all'OED Appeals sul web a seguito del normale e rigoroso processo di revisione da parte dell'editore.

Altro esempio è il Collins, che però conta di un diverso processo di crowdsourcing: 1. gli utenti inviano i loro suggerimenti per le nuove entrate lessicali attraverso il CollinsDictionary.com. Inoltre, la casa editrice ha recentemente lanciato la

---

<sup>143</sup> <http://public.languagedepot.org/>

<sup>144</sup> <http://wesay.palaso.org/about-us/project-summary/>

<sup>145</sup> <http://public.oed.com/the-oed-today/recent-updates-to-the-oed/march-2014-update/new-words-list-march>

campagna #twictionary, cioè gli utenti di Twitter sono invitati a votare su quali tra molte parole vorrebbero vedere incluse nella successiva edizione del Collins English Dictionary; 2. i lessicografi del Collins valutano che le entrate lessicali, suggerite sulla base del Collins Corpus, siano abbastanza diffuse per essere incluse nel dizionario; 3. le entrate lessicali *crowdsourced* e le loro definizioni sono aggiornate una volta al mese con i suggerimenti dei migliori utenti. Gli username delle persone che hanno suggerito le parole approvate vengono elencati nelle pagine delle entrate del dizionario come riconoscimento per il loro contributo.

3. La raccolta di risorse linguistiche di crowdsourcing attraverso piattaforme **collaborative** rappresenta una delle opzioni veloci e a costi moderati nella ricerca di dati di informazione linguistica dal crowd sia per dizionari convenzionali sia per dizionari *machine readable*, cioè dizionari annotati per compiti di NLP: Word Sense Disambiguation (WSA), Sentiment analysis, traduzione automatica e assistita.

Mechanical Turk (Mturk) è una delle piattaforme che fornisce un sistema di servizi web offerto da Amazon.com, il quale utilizza il contributo degli utenti per lo svolgimento di compiti distribuiti su larga scala. I *requesters* postano Human Intelligence Tasks (HITs) da completare dai cosiddetti *turkers* (gli operatori) dietro compenso. I *turkers* scelgono i loro HITs, completano il lavoro (ad es. traduzione di parole<sup>146</sup>, creazione di lessici, annotazione di dizionari) inviano i risultati. Nell'organizzazione del sistema gli utenti che vogliono accedere sia come *requester* sia come *turker* hanno bisogno di avere un account Amazon; gli account Amazon sono anonimi, ma hanno un unico ID Amazon. I *requesters* possono impostare un HIT specificando il compito da svolgere e il compenso previsto. Una volta immessa la richiesta, i *requesters* possono limitare l'accesso all'HIT ai soli lavoratori con particolari qualifiche, come sufficiente precisione in un test o una minima percentuale di *submissions* accettate in precedenza. Dopo che ogni HIT viene completato, il *requester* può approvare il lavoro e compensare il *turker* attraverso Amazon ed eventualmente scegliere di dare un bonus.

---

<sup>146</sup> E. PAVLICK, M. POST, A. IRVINE, D. KACHAEV, C. CALLISON-BURCH, *The Language Demographics of Amazon Mechanical Turk*, "Transactions of the Association for Computational Linguistics", 2 (Feb), pp. 79-92.

### III. 2 Nooj<sup>147</sup>

L'applicazione Nooj è stata scelta nel presente lavoro di tesi per dimostrare come sia necessario un complesso lavoro di documentazione delle risorse linguistiche affinché possano essere disponibili per applicazioni informatiche come ad esempio NLP, traduzione automatica e assistita e costruzione di grammatiche flessive. Tra i programmi elencati nei precedenti paragrafi è stato scelto Nooj perché rispetto agli altri programmi ha delle funzionalità di portabilità più avanzate, e anche il livello di intuizione è più semplice e può coinvolgere una fascia più vasta di utenti.

Nooj è un programma per il trattamento automatico del linguaggio naturale il cui autore è Max Silberztein. Nooj viene utilizzato per le sue funzionalità di interrogazione di corpora (ad es. per analisi di testi letterari, ricerca ed estrazione di informazione da giornali o corpora tecnici), per formalizzare alcuni fenomeni linguistici (ad es. descrizione della morfologia di una lingua), e per applicazioni computazionali (analisi di testo automatica)<sup>148</sup>.

Nooj rappresenta l'evoluzione di INTEX (a cui Silberztein lavorò nel decennio 1992-2002 sotto la guida del linguista Maurice Gross). Per costruire e completare la nuova piattaforma l'autore ha usato la propria esperienza decennale come progettista, programmatore e utente di INTEX. Nel corso degli anni la tecnologia INTEX era diventata obsoleta: era monolingue, poteva trattare solo un singolo formato file, un file di testo alla volta e non supportava file XML<sup>149</sup>.

Con Nooj è possibile costruire dizionari elettronici e grammatiche e applicarli a corpora di grandi dimensioni grazie a un motore linguistico che include dispositivi computazionali utilizzati sia per la descrizione formale dei fenomeni linguistici sia per il *parsing* dei testi scritti: automi e traduttori a stati finiti, reti ricorsive, espressioni regolari e grammatiche *context-free*. Tramite Nooj si possono costruire grammatiche anche bilingui da associare ai relativi dizionari<sup>150</sup>. Queste sono usate per rappresentare una vasta gamma di fenomeni linguistici, dai livelli ortografici e morfologici a quelli sintattici e trasformativi. In particolare:

---

<sup>147</sup> Per ulteriori approfondimenti cfr. Nooj Manual

<sup>148</sup> S. VIETRI, op. cit., pp. 13-14-15.

<sup>149</sup> Per ulteriori approfondimenti cfr. Nooj Manual

<sup>150</sup> All'interno di Nooj sono stati costruiti, da diverse equipe di ricerca, moduli relativi a lingue come l'inglese, il francese, l'arabo, l'ungherese, il portoghese, l'armeno, il bulgaro, il catalano, il cinese, l'ebraico, il latino, lo spagnolo, l'accadico, il polacco e l'italiano.

- le grammatiche flessive e derivazionali (.nof file) vengono usate per rappresentare proprietà flessive (ad es. la coniugazione) o derivazionali (ad es. nominalizzazione) delle entrate lessicali. Queste descrizioni possono essere immesse sia graficamente che in forma di regole;
- le grammatiche lessicali, ortografiche, morfologiche o terminologiche (.nom file) sono usate per rappresentare insiemi di forme di parole, e associarle con informazioni lessicali, ad es. per standardizzare lo spelling di una parola o la variante di una parola, per riconoscere e taggare neologismi, per collegare espressioni sinonimiche;
- le grammatiche sintattiche, semantiche e di traduzione (.nog file) sono usate per riconoscere e annotare espressioni nei testi, ad es. per taggare gruppi nominali, alcuni costrutti sintattici o espressioni idiomatiche, per estrarre alcune espressioni o interessi (nomi di società, espressioni di date, indirizzi ecc), per disambiguare le parole filtrando alcune annotazioni lessicali o sintattiche nel testo per effettuare trasformazioni (ad es. parafrasare automaticamente) o traduzione automatica (da una lingua di partenza a un'altra).

Le funzioni di Nooj si trovano nella noojapply.exe, disponibile per il download; noojapply.exe permette agli utenti di compilare un dizionario, e applicare ai testi dizionari e grammatiche in modo automatico.

### III. 2. 1 I dizionari Nooj<sup>151</sup>

I dizionari Nooj hanno superato i dizionari INTEX perché il nuovo sistema permette la costruzione di dizionari elettronici sia di parole semplici sia di parole composte in forma canonica e flessa. È possibile costruire in modo autonomo dizionari per la traduzione automatica, importare testi in più di 100 formati e costruire corpora di grandi dimensioni<sup>152</sup>.

---

<sup>151</sup> Per ulteriori approfondimenti cfr. Nooj Manual.

<sup>152</sup> G. DE BUERIIS ET AL., op.cit., 2005, p. 14.

Il primo livello di analisi del testo richiede che il computer identifichi le *Atomic Linguistic Units (ALUs)* del testo, cioè i suoi elementi non analizzabili. Questi sono costituiti da parole il cui significato non può essere *predicted*, perciò è importante descriverle in modo esplicito. Da un punto di vista formale Nooj separa le ALUs in 4 classi: SIMPLEWORDS; AFFIXES; MULTI-WORD UNITS; FROZEN EXPRESSIONS.

Nei dizionari Nooj tutte le entrate sono associate a una categoria morfo-sintattica. Possiamo così fare riferimento a questa categoria nelle espressioni regolari: ad esempio, AGG aggettivo, AVV avverbio, CONG congiunzione e così via. Ad ogni modo, i linguisti e i lessicografi che usano Nooj sono totalmente liberi di inventare le loro categorie e codici. Gli utenti infatti possono aggiungere il proprio *code system*, sia in nuovi dizionari personali che, modificando i dizionari del sistema, possono essere istantaneamente usati in qualsiasi espressione regolare o grammatica.

Le informazioni (inclusi i codici flessivi) in Nooj sono rappresentate da codici preceduti dal carattere + . I codici flessivi non sono di solito visibili nei dizionari Nooj, perché questi contengono lemmi piuttosto che forme coniugate. È importante sapere quali siano questi codici perché possono essere usati esattamente come codici sintattici o semantici. Le entrate lessicali di Nooj possono inoltre essere collegate a un superlemma che si comporta come forma canonica per le entrate lessicali così come tutte le sue forme derivate e flesse. Il modulo della flessione in Nooj viene azionato aggiungendo la speciale proprietà +FLX all'entrata lessicale.

Nooj fornisce due equivalenti strumenti per descrivere questi paradigmi flessivi: sia graficamente sia tramite regole testuali. Entrambe le descrizioni sono equivalenti, e internamente compilate in un *Finite-State Trasducer* (in .nof file). Il dizionario deve contenere almeno una riga che inizia con il comando:

```
#use NameOfAnInflectionalDescriptionFile.nof
```

I nomi del paradigma nel caso della descrizione della flessione in modo testuale corrispondono alle regole incluse nel file della descrizione flessiva.

Es: help ,V+FLX=ASK

Deve esistere nel file di descrizione .nof una regola chiamata ASK che descrive tutte le forme coniugate del verbo *to help*. Tutti i verbi che coniugano in questa maniera sono associati alla stessa regola.

Nooj utilizza i seguenti operatori che permettono di costruire ogni tipo di paradigma di flessione:

- <E> stringa vuota
- <B> cancella l'ultimo carattere
- <D> duplica l'ultimo carattere
- <L> va a sinistra
- <R> va a destra
- <N> va alla fine della parola successiva
- <P> va alla fine della parola precedente
- <S> cancella il carattere successivo

È anche possibile definire la relazione tra categorie, proprietà e i loro valori usando un *Dictionary Properties' Definition file*. In questo file gli utenti possono descrivere quali proprietà sono rilevanti per ogni categoria morfosintattica, e quali valori sono aspettati per ogni proprietà. Ad es:

```
V_Pers = 1 | 2 | 3
V_Nb = s | p
V_Tense = G | INF | PP | PR | PRT
V_Syntax = aux | i | t
```

Alcuni valori delle proprietà possono essere caratteristiche flessive:

```
INFLECTION = 1 | 2 | 3 | m | f | n
             | s | p | G | INF | PP | PR | PRT
```

Per ogni lingua Nooj accede a un dizionario nel quale ogni parola della lingua è un'entrata, ed è associata all'informazione morfologica, di solito il suo paradigma

flessivo e/o derivazionale. Il paradigma flessivo comunica a Nooj quali forme flesse sono accettate dall'entrata lessicale, cioè quali sono le sue forme coniugate (se si tratta di un verbo), le sue forme femminili e plurali (per nomi in lingue romanze), le sue forme in accusativo, dativo, genitivo ecc. (ad es. per le lingue germaniche). Grazie a questo dizionario, Nooj può collegare tutte le forme flesse insieme. Si può accedere a queste informazioni semplicemente immettendo qualsiasi lemma tra virgolette uncinato. Ad es. la seguente espressione nella quale ci riferiamo alla forma della parola *be*, rappresenta tutte le forme flesse di *to be*, alla quale siamo interessati.

<be>

Per analizzare i testi, Nooj ha bisogno di dizionari che raccolgano e descrivano tutte le parole di un testo, così come alcuni meccanismi da collegare a queste entrate lessicali alle corrispondenti (flesse e/o derivate) forme che realmente occorrono nel testo.



### III. 3 Il dizionario gallurese

Il dizionario gallurese è stato ideato sulla base di dizionari cartacei già esistenti, implementati poi da conoscenze personali e ricerche ulteriori. I dizionari scelti sono stati il *Dizionario della lingua gallurese* di Francesco Rosso e il *Vocabolario tempiese-italiano, italiano-tempiese* di Andrea Usai. La scelta è ricaduta su due dizionari poiché entrambi incompleti dal punto di vista etimologico. Per quanto riguarda invece la stesura della grammatica il lavoro si basa sulla “grammatichetta” che si trova nella prima parte del dizionario dell’Usai. La scelta è stata dettata anche da un rapporto di interscambio con la Consulta del Gallurese che è stata intervistata durante la stesura del presente lavoro.

È stato costruito un prototipo di lessemi della lettera A, con relativi tag. Il dizionario contiene 5172 entrate non taggate interamente per le tempistiche che richiederebbe la costruzione di un dizionario-macchina di tale portata, di seguito un esempio:

```
abbaccàtu,AGG+Gen=m+Num=s+FLX=ABBACCÀTU+GS+IT="affossa  
to"  
abbaccàtu,AGG+Gen=m+Num=s+FLX=ABBACCÀTU+GS+IT="abbassa  
to"  
abbacchjà,V+tr+T=INF+FLX=ABBÀ+GS+IT="abbacchiare"  
abbacchjà,V+tr+T=INF+FLX=ABBÀ+GS+IT="mortificare"  
abbacchjatu,AGG+Gen=m+Num=s+FLX=ABBACCÀTU+GS+IT="abbac  
chiato"  
abbacchjatu,AGG+Gen=m+Num=s+FLX=ABBACCÀTU+GS+IT="morti  
ficato"  
abbàccu,N+Genere=m+Num=s+FLX=ABBÀCCU+GS+IT="scavo"  
abbaccu,N+Genere=m+Num=s+FLX=ABBÀCCU+GS+IT="avvallamen  
to"  
abbadalchjà,V+tr+T=INF+FLX=ABBADALCHJÀ+GS+IT="abbatacc  
hiare"  
abbaddhà,V+tr+T=INF+FLX=ABBADALCHJÀ+GS+IT="Il  
bloccarsi della pallottola in canna"
```

L’ortografia utilizzata è doppia, poiché si è deciso di integrare quella poetica e quella comune, che si è sviluppata negli ultimi anni. In genere i due tipi di ortografia differiscono per la presenza o meno di accenti, o la presenza o meno della lettera *h* dopo le doppie *gg dd e cc*.

Il lavoro di analisi testuale e costruzione del dizionario macchina con relativa flessione e grammatica è stato invece svolto su un verso di una poesia inedita:

La linga è tarra,  
disiu di esse illu mundu.  
La linga è focu,  
paldutu lu caminu  
poni luci illu cori.

La poesia è stato poi ridotta in lemmi che sono stati taggati e riportati nella morfologia flessiva:

la, DET+Gen=f+Num=s+FLX=LA+GS+IT="la"  
linga, N+Gen=f+Num=s+FLX=LINGA+GS+IT="lingua"  
esse, V+t+intr+INF+FLX=ESSE+GS+IT="essere"  
tarra, N+Gen=f+Num=s+FLX=TARRA+GS+IT="terra"  
disiu, N+Gen=m+Num=s+FLX=DISIU+GS+IT="desiderio"  
di, PREP+GS+IT="di"  
illu, PREP+Gen=m+Num=s+FLX=ILLU+GS+IT="nel"  
mundu, N+Gen=m+Num=s+FLX=DISIU+GS+IT="mondo"  
focu, N+Gen=m+Num=s+FLX=FOCU+GS+IT="fuoco"  
paldutu, AGG+Gen=m+Num=s+FLX=PALDUTU+GS+IT="perduto"  
lu, DET+Gen=m+Num=s+FLX=LU+GS+IT="il"  
caminu, N+Gen=m+Num=s+FLX=DISIU+GS+IT="cammino"  
punì, V+t+INF+FLX=PUNÌ+GS+IT="mettere"  
luci, N+Gen=f+Num=s+FLX=LUCI+GS+IT="luce"  
cori, N+Gen=m+Num=s+FLX=CORI+GS+IT="cuore"

Ai lessemi taggati è stata poi assegnata una grammatica e le relative properties' definition, es.:

#VERBS#

AÉ

<E>/INF | <B>utu/PP | <B>enti/PPR | <B>endi/GER

IND

| <B>gghju/PRE+1+s | <B>i/PRE+2+s | <B>a/PRE+3+s |  
<B>emu/PRE+1+p | <B>eti/PRE+2+p | <B>ni/PRE+3+p |

PAR

| <B>ési/PAR+1+s | <B>esti/PAR+2+s | <B>esi/PAR+3+s |  
<B>ésimi/PAR+1+p | <B>ésiti/PAR+2+p | <B>ésini/PAR+3+p |

IMF

| <B>ia/IMF+1+3+s | <B>ii/IMF+2+s | <B>iami/IMF+1+p |  
<B>iati/IMF+2+p | <B>iani/IMF+3+p |

FUT

| <B>ràggju/FUT+1+s | <B>ré/FUT+2+s | <B>rà/FUT+3+s |  
<B>remu/FUT+1+p | <B>réti/FUT+2+p | <B>rani/FUT+3+p |

Il funzionamento di Nooj verrà dimostrato in sede di discussione di laurea, in cui verrà mostrato un esempio di *inflection* a partire dalla poesia sopra citata.

### III. 4 Ulteriori considerazioni

In merito alla tesi esposta è possibile aggiungere come ultima analisi conclusiva che in materia di apprendimento di una lingua che sta lentamente cadendo in disuso è fondamentale l'apprendimento facilitato di tale idioma. Al momento il mondo dell'eLearning è in costante espansione. Gli utenti possono provvedere alla propria formazione per ciò che concerne qualsiasi materia direttamente online. È il caso ad esempio delle piattaforme Moodle, in cui vengono condivisi contenuti materia di esame nei siti delle università, o è il caso ad esempio delle università telematiche. Un'ottica di apprendimento in questo senso potrebbe essere vantaggiosa e costruttiva anche in materia di didattica di dialetto e lingue minoritarie. Il contributo degli utenti e dei docenti esperti della materia linguistica in congiunzione potrebbero dar vita a veri e propri corsi di lingua di un livello sempre più alto.

## *Conclusioni*

Il progetto esposto ha voluto dare validità legislativa e dignità ad una varietà linguistica come parte della biodiversità delle lingue mondiali. Il gallurese è quindi una lingua che va tutelata esattamente come tutte le altre lingue globali.

Si è voluto fornire un mezzo per avviare il progetto di tutela in direzione delle nuove scoperte informatiche. Nella ricerca dell'elaborato esposto è stata fondamentale la documentazione reperita online, questo dimostra quanto le precedenti teorie e ricerche in ambito di salvaguardia delle lingue minoritarie e costruzione di mezzi di tutela vada in direzione dell'informatica e non della salvaguardia delle varietà linguistiche di vecchio stampo.

Il crowdsourcing e il cloudcomputing sono i nuovi mezzi di ricerca di informazioni e di cooperazione in ambito sia linguistico che di studi in generale. Non è più possibile per la ricerca scindere dalla comunicazione online e da programmi di connessione tra utenti che immagazzinano e favoriscano lo scambio simultaneo di informazioni.

Il presente progetto è stato applicato alla lingua gallurese, ma può essere esteso in ambito più ampio, sulla tutela delle lingue minoritarie in generale, ma anche su progetti di più ampio respiro come l'interconnessione informativa tra utenti che ricercano in ogni campo.

La tesi si conclude con una riflessione di quello che è stato il mio percorso di ricerca e di scrittura della stessa. La ricerca del materiale e i mezzi necessari alla costruzione di un dizionario informatizzato hanno evidenziato quanto sia scarna la ricerca in merito di valorizzazione delle varietà sarde e delle lingue minoritarie in generale, come più volte ribadito nell'elaborato il crowdsourcing, il cloudcomputing e il NLP possono fare molto in tema di raccolta, elaborazione e analisi di dati linguistici. La stesura di un dizionario informatizzato con mezzo di cloudcomputing con un mezzo come Nooj potrebbe essere un trampolino di lancio anche per altre varietà sarde e per le lingue minoritarie.

## Bibliografia

- ABEL, C. M. MEYER, *The dynamics outside the paper: user contribution to online dictionaries* in Proceedings of the 3rd Biennial Conference on Electronic Lexicography (eLex), Ljubljana, Trojina, Institute for Applied Slovene Studies/Tallinn, Eesti Keele Instituut, Tallinn, Estonia, October 2013;
- ABLEY M., *Spoken Here*, Arrow Books, Berkshire (UK), 2003;
- ADURIZ I. ET AL., *EUSLEM: A lemmatiser/tagger for Basque* in Proc. Of EURALEX '96, 1996;
- ALEGRIA I. ET AL., *Robustness and customisation in an analyser/lemmatiser for Basque* in Proceedings of Workshop on Customizing knowledge in NLP applications, Third International Conference on Language Resources and Evaluation, 2002;
- ANGIUV. S in G. CASALIS, *Dizionario geografico storico statistico degli Stati di S. M. il Re di Sardegna*, vol. VII, Torino, voce Gallura, 1840;
- ASSEMBLEA DELLA CONSULTA, *Comunicato dell'Assemblea*, sede della consulta del gallurese, Arzachena, 2001;
- ATZORI M., SATTA M. M., *Credenze e riti magici in Sardegna*, Sassari, Chiarella, 1980;
- BARCELLONA N., MARINI, MONTI, VERCESI, *5000 termini dell'informatica*, Milano, Gruppo Editoriale Jackson, 1988;
- BECCARIA G. L. (a cura di), *Dizionario di linguistica*, Bologna, 1984;
- BELLODI W., *Ilienses, Balari e Corsi antiche popolazioni della Sardegna*, in *Sardegna Mediterranea*, n. 20, 2006;

- BELLU G., rel. CLARA GALLINI, tesi di laurea: *Storia religiosa gallurese: alcuni aspetti*, Università di Cagliari, a. a. 1970;
- BLASCO FERRER E., *Storia linguistica della Sardegna*, Tubingen, Max Niemeyer, 1984;
- BRAUDEL F. , *Civiltà e imperi del Mediterraneo nell'età di Filippo II*, Torino, Piccola Biblioteca Einaudi, 2010;
- CASTANO S. ET AL., *Informazione, conoscenza e Web per le scienze umanistiche*, Pearson Addison Wesley, 2009;
- CONSIGLIO D'EUROPA, *Carta Europea delle lingue regionali o minoritarie*, , Preambolo, 1992;
- CHIARI I., *Introduzione alla linguistica computazionale*, Nota dell'autrice, Bari, Editori Laterza, 2007;
- CHOMSKY N., *The Logical Structure of Linguistic Theory*, 1975;
- COMMISSIONE EUROPEA, *Parlare le lingue europee, Le lingue nell'Unione europea*, ec.europa.eu/publications, Direzione generale della Comunicazione, Pubblicazioni, Lussemburgo, Ufficio delle pubblicazioni ufficiali delle Comunità europee, 2008;
- CONSULTA INTERCOMUNALE GALLURESE , *Documento sulla lingua sarda unificata*, Arzachena, 2003;
- CONSULTA INTERCOMUNALE GALLURESE, *Gallura. Cenni storici e diversità linguistiche*, Olbia, Editrice Taphros, 2003;
- CORDA F., *Saggio di grammatica gallurese*, Cagliari, Editrice 3T, 1983;

- DE BOFARULL I MASCARÓ P., *Colección de documentos inéditos del Archivo de la Corona de Aragón*, XI, S.I., Establecimiento Litográfico y Tipográfico de José Eusebio Monfort, 1847-1851;
- DE BUERIIS G. ET AL., *Lessici elettronici e descrizioni lessicali, sintattiche, morfologiche ed ortografiche*, Salerno, Plectica Editrice, 2005;
- DE MARTINO R., *Il Dialetto Gallurese* in a cura di SALVATORE BRANDANU, *La Gallura, una Regione diversa in Sardegna*, San Teodoro, Editrice I.C.I. Mar, 2001;
- DEIDDA G., A. DELLA MARIA, *Sagre, riti e feste popolari in Sardegna*, Cagliari, Janus, 1987;
- EL-HAJ M. ET AL., *Creating language resources for under-resourced languages: methodologies, and experiments with Arabic in Language Resources and Evaluation*, 2015;
- FARA F., *Chorographia Sardiniae (1580-85), Opera*, Vol I, a cura di E. CADONI, Sassari, 1992;
- FUERTES OLIVERA P. A., *The Function Theory of Lexicography and Electronic Dictionaries: WIKTIONARY as a Prototype of Collective Free Multiple-Language Internet Lexicography*, H. Bergenholtz, S. Nielsen & S. Tarp (eds.) *Lexicography at a Crossroads: Dictionaries and Encyclopedias Today, Lexicographical Tools Tomorrow*, Bern, Peter Lang, 2009;
- GRISHMAN R., *Linguistica computazionale*, Milano, Tecniche Nuove, 1988;
- GROSS M., *Local Grammars and their Representation by Finite Automata*, in Hoey M. (ed.), *Data, Description, Discourse. Papers on the English language in honour of John McH Sinclair*, Harpers Collins, London, 1993;



- HARRIS Z. S., *From Morpheme to Utterance*, in *Language*, n. 22, 1946;
- JONES M.P., *Lingue a rischio di estinzione e diversità linguistica nell'Unione Europea*, Parlamento Europeo, 2013;
- L. N. 482/1999, Interventi nel settore dei mezzi di comunicazione, internet e nuove tecnologie, Capo IV;
- L. R. 26/97, Titolo I (principi e finalità), articolo 2 oggetto, punto 1, *Promozione e valorizzazione della cultura e della lingua della Sardegna*;
- LE LANNOU M., *Pâtres et paysans de la Sardaigne* in *Revue de géographie alpine*, Volume 30, Numero 3, 1942;
- LE VIET BAC ET AL., *Using the web for fast language model construction in minority languages* in *INTERSPEECH*, 2003;
- LEW R., *Online dictionaries of English*, P.A. Fuertes-Olivera & H. Bergenholtz (eds.) *e-Lexicography: The Internet, Digital Initiatives and Lexicography*, London/New York, Continuum, 2011;
- LEW R., *User-generated content (UGC) in English online dictionaries*, A. Abel & A. Klosa (eds.) *Ihr Beitrag bitte! – Der Nutzerbeitrag im Wörterbuchprozess* (OPAL – Online publizierte Arbeiten zur Linguistik), Institut für Deutsche Sprache, Mannheim, 2013;
- MAXIA M., *I Corsi in Sardegna*, Cagliari, Edizioni Della Torre, 2006;
- MAXIA M., *Studi storici sui dialetti della Sardegna settentrionale*, Sassari, Stadium Adf, 1999;
- MILLARDET G., *Sur un ancien substrat commun à la Sicilie, la Corse, la Sardaigne*, in *Revue delinguistique romane*, Vol. 9, 1933;

- MONTI J., *Le tecnologie per la traduzione nell'era del cloud computing: stato dell'arte e prospettive future* in *Atti del III workshop internazionale Traduzione per le aziende, il territorio, l'editoria: teoria, prassi, didattica*, Centro Linguistico di Ateneo Università di Salerno, 18-20 dicembre 2013;
- MORVAN P., *Dizionario di informatica*, Roma, Gremese-Larousse, 1989;
- PANU T., *Storia di Tempio e della Gallura*, Muros (SS), Nuova Stampa Color, 2010;
- PAVLICK E., M. POST, A, IRVINE, D, KACHAEV, C. CALLISON-BURCH, *The Language Demographics of Amazon Mechanical Turk*, *Transactions of the Association for Computational Linguistics*, s.d.;
- PETNAKOV I., *Appunti sui dialetti sardo-settentrionali*, in *Archivum Romanicum*, XXV, 1921;
- POST M., CALLISON-BURCH C., OSBORNE M., *Constructing parallel corpora for six indian languages via crowdsourcing* in: *Proceedings of the Seventh Workshop on Statistical Machine Translation*. Association for Computational Linguistics, 2012;
- RATTU S., *S. Teresa di Gallura*, in *Studi storici in onore di F. Loddo Canepa*, Bologna, 1956, vol. I;
- REGIONE AUTONOMA DELLA SARDEGNA, *Relazione di accompagnamento al D.D.L. in materia di Norme per la tutela, valorizzazione e promozione della lingua sarda e delle altre varietà linguistiche della Sardegna*, 1997;
- ROHLFS G., *Grammatica storica della lingua italiana e dei suoi dialetti*, Torino, Einaudi, 1969;

- ROSENBAUM P. S., *The Grammar of English Complement Constructions*, MIT Press, Cambridge, 1967;
- ROSSO F., DEBIDDA A., FRESI L., *Dizionario della lingua gallurese*, Tempio, StampaSi, 2000;
- TOSO F., *Il Patrimonio linguistico sardo* Parte VIII – Aspetti del panorama linguistico europeo, in *LINGUE E DIALETTI*, cap. 20;
- TOSO F. *Patrimoni linguistici e lingue minoritarie: la prospettiva europea e quella italiana* in *Annali della Facoltà di Lingue e Letterature Straniere dell'Università di Sassari*, 2009, Vol. 5, Articolo. 119;
- TOSO F., *La Sardegna che non parla sardo*, Cagliari, CUEC Editrice, 2012;
- TURTAS R., *Le chiese nel verde*, Sassari, 1988;
- UNIONE EUROPEA, *Carta dei diritti fondamentali dell'Unione Europea*, 2000;
- USAI A., *Vocabolario tempiese-italiano, italiano-tempiese*, Sassari, Ed. Poddighe, 1977;
- VIETRI S., *Dizionari elettronici e grammatiche a stati finiti*, Salerno, Plectica editrice, 2008;
- WAGNER M. L., a cura di G. PAULIS, *La lingua sarda*, Nuoro, Ilisso Edizioni, 1997;
- WAGNER M. L., *Fonetica del sardo* (Introd, trad. e appendice di G. PAULIS), Cagliari, Gianni Trois Editore, 1984.

## *Sitografia*

- [creativecommons.org/licenses/by-sa/3.0/it/](https://creativecommons.org/licenses/by-sa/3.0/it/);
- [demo.taas-project.eu/](https://demo.taas-project.eu/);
- [en.wiktionary.org/wiki/Wiktionary:Main\\_Page](https://en.wiktionary.org/wiki/Wiktionary:Main_Page);
- [en.wiktionary.org/wiki/Wiktionary](https://en.wiktionary.org/wiki/Wiktionary);
- [gov.wales/topics/welshlanguage/promoting/capital-investment-fund-2015-to-2016/?lang=en](https://gov.wales/topics/welshlanguage/promoting/capital-investment-fund-2015-to-2016/?lang=en);
- [igm.univ-mlv.fr/%7Eunitex/](https://igm.univ-mlv.fr/%7Eunitex/);
- [info.melt-project.eu/ww/en/pub/melt\\_project/welcome.htm](https://info.melt-project.eu/ww/en/pub/melt_project/welcome.htm);
- [intex.univ-fcomte.fr/](https://intex.univ-fcomte.fr/);
- [it.termwiki.com/](https://it.termwiki.com/);
- [lingro.com/](https://lingro.com/);
- [mcmillandictionary.com/open-dictionary/latestEntries.htm](https://mcmillandictionary.com/open-dictionary/latestEntries.htm);
- [public.languagedepot.org/](https://public.languagedepot.org/);
- [public.oed.com/appeals/](https://public.oed.com/appeals/);
- [public.oed.com/the-oed-today/recent-updates-to-the-oed/march-2014-update/new-words-list-march](https://public.oed.com/the-oed-today/recent-updates-to-the-oed/march-2014-update/new-words-list-march);

- [wesay.palaso.org/about-us/project-summary/](http://wesay.palaso.org/about-us/project-summary/);
- [wikimediafoundation.org/wiki/Pagina\\_Principale](http://wikimediafoundation.org/wiki/Pagina_Principale);
- [www.europarl.europa.eu/charter/pdf/text\\_it.Pdf](http://www.europarl.europa.eu/charter/pdf/text_it.Pdf)
- [www.coe.int/en/web/conventions/full-list/-/conventions/treaty/157](http://www.coe.int/en/web/conventions/full-list/-/conventions/treaty/157);
- [www.collinsdictionary.com/cas/login?service=http%3A%2F%2Fwww.collinsdictionary.com%2Fj\\_spring\\_cas\\_security\\_check](http://www.collinsdictionary.com/cas/login?service=http%3A%2F%2Fwww.collinsdictionary.com%2Fj_spring_cas_security_check);
- [www.consultaintercomunalegallura.net/](http://www.consultaintercomunalegallura.net/);
- [www.cstr.ed.ac.uk/~briony/SALTIMIL/](http://www.cstr.ed.ac.uk/~briony/SALTIMIL/);
- [www.gnu.org/gnu/thegnuproject.it.html](http://www.gnu.org/gnu/thegnuproject.it.html);
- [www.gov.im/](http://www.gov.im/);
- [www.isca-speech.org.](http://www.isca-speech.org;);
- [www.mturk.com/mturk/welcome](http://www.mturk.com/mturk/welcome);
- [www.nooj4nlp.net/](http://www.nooj4nlp.net/);
- [www.per.gov.ie/en/20-year-strategy-for-the-irish-language-2010-2030/](http://www.per.gov.ie/en/20-year-strategy-for-the-irish-language-2010-2030/);
- [www.regjeringen.no/globalassets/upload/fad/vedlegg/sami/hp\\_2009\\_samisk\\_spraak\\_engelsk.pdf](http://www.regjeringen.no/globalassets/upload/fad/vedlegg/sami/hp_2009_samisk_spraak_engelsk.pdf);
- [www.unesco.org/languages-atlas/](http://www.unesco.org/languages-atlas/);
- [www.urbandictionary.com/](http://www.urbandictionary.com/);

- [www.uzei.eus/](http://www.uzei.eus/);
- [xuxen.eus/](http://xuxen.eus/);
- [www.regione.sardegna.it/](http://www.regione.sardegna.it/).