



Unione europea
Fondo sociale europeo



Repubblica Italiana



REGIONE AUTONOMA DELLA SARDEGNA



obiettivo competitività regionale e occupazione

ASSESSORATO DELLA PUBBLICA ISTRUZIONE, BENI CULTURALI, INFORMAZIONE,
SPETTACOLO E SPORT

**Direzione Generale della Pubblica Istruzione
Servizio Istruzione**

POR FSE Sardegna 2007 – 2013

semid@s

**GARA CON PROCEDURA APERTA PER L'APPALTO DI SERVIZI
OPERAZIONE SCUOLA DIGITALE - REALIZZAZIONE DEL
SISTEMA TELEMATICO, PRODUZIONE DI CONTENUTI DIDATTICI
DIGITALI E COSTITUZIONE DI UN CENTRO DI COMPETENZE PER
L'EROGAZIONE DI SERVIZI DI ECCELLENZA**

ALLEGATO VI: LINEE GUIDA SULLA LINGUA SARDA



Unione europea
Fondo sociale europeo



Repubblica Italiana



REGIONE AUTONOMA DELLA SARDEGNA



obiettivo competitività regionale e occupazione

L'Assessorato della Pubblica Istruzione, Beni Culturali, Informazione, Spettacolo e Sport ritiene importante allargare l'esperienza della creazione dei contenuti multimediali del progetto Scuola Digitale a una sperimentazione importante sul versante dell'utilizzo della Lingua e a una valorizzazione più propria di alcuni elementi della Cultura Sarda. In particolare, si ritiene di aderire, quale documento fondamentale, ai principi ispiratori e ai contenuti di programma del Piano Triennale della Lingua e Cultura Sarda 2011-2013 approvato recentemente con delibera di Giunta Regionale n°46/47 del 16/11/ 2011.

Nel piano sono elencate le linee fondamentali che ispirano le azioni del Governo Regionale nell'azione di contrasto al processo di depauperazione e sostituzione della lingua sarda (e delle lingue alloglotte presenti in Sardegna) nell'ambito di una strategia complessiva che mira a sostenere un'ampia sensibilizzazione sociale nella direzione del rafforzamento di un multilinguismo effettivo e pregnante della società sarda e delle sue istituzioni scolastiche.

La questione della difesa della differenza e varietà linguistica della Sardegna va pertanto inquadrata nell'ambito educativo della necessità di affermare modelli didattici che riconoscano nel multilinguismo l'obiettivo primario da raggiungere all'interno del mondo nella scuola pur nel rispetto dell'autonomia di ogni singola istituzione che sia situata nell'isola.

L'orizzonte giuridico di tale intervento risiede innanzitutto nell'articolo 6 della Costituzione repubblicana e nella legge statale 482/99 che ha riconosciuto e normato con rigore l'utilizzo di 12 lingue minoritarie, tra cui in Sardegna quella sarda e catalana, all'interno delle scuole di ogni ordine e grado. Tali previsioni giuridiche si sono incrociate negli anni con quelle applicate in Sardegna in seguito all'approvazione della legge regionale n°26 del 1997 e all'applicazione dei piani triennali linguistico-culturali della stessa Regione Autonoma della Sardegna. In particolare, a partire dal Piano Triennale 2008-2010, la Regione ha messo l'accento in maniera più rilevante e cogente, sulla salvaguardia linguistica delle varietà presenti nel territorio isolano con una serie di misure che hanno trovato ulteriore realizzazione e migliore applicazione nel Piano Triennale 2011-2013 citato sopra.

I principi fondamentali a cui l'azione di promozione della diversità linguistica della Sardegna si deve improntare sono i seguenti:

- a) salvaguardia del patrimonio immateriale linguistico della Sardegna attraverso il riconoscimento dei diritti linguistici dei parlanti sardo o lingue alloglotte;
- b) contrasto al processo di sostituzione linguistica della lingua sarda e di quelle alloglotte con azioni tese a rafforzarne il prestigio sociale e la trasmissione intergenerazionale;
- c) rafforzamento della didattica della lingua minoritaria e suo radicamento nel mondo della scuola attraverso l'implementazione di curricula formativi da consolidare con corsi universitari e privati, borse di studio, campagne di sensibilizzazione e promozione;
- d) contrasto alla ghetizzazione della lingua, alla sua stigmatizzazione sociale e al suo uso esclusivo nella veicolazione della cultura locale, tradizionale e rurale;
- e) sostegno e formazione costante degli operatori della politica linguistica;
- f) l'implementazione delle norme linguistiche di riferimento a carattere sperimentale per la lingua sarda scritta in uscita dell'Amministrazione regionale e il completamento di una norma ortografica per tutte le varietà dell'isola;
- g) consolidamento del patrimonio dialettologico di tutta l'isola;
- h) decentramento armonico delle competenze agli enti locali;
- i) rafforzamento della presenza della lingua minoritaria sui media regionali;
- j) rafforzamento della presenza della lingua minoritaria nella pubblica amministrazione;
- k) sostegno alla rivitalizzazione della lingua in campo liturgico ed ecclesiale;
- l) sostegno alla presenza della lingua sarda nei settori artistici della letteratura;
- m) negazione del fatto che la questione della lingua venga ridotta a "nicchia" specialistica di studiosi ed esperti e torni a essere un fatto veramente popolare e di interesse generale.

In particolare, per la buona riuscita delle politiche linguistiche di rivitalizzazione di una lingua minoritaria quale quella sarda, è sembrato fondamentale ai pianificatori linguistici spezzare il nesso quasi obbligato (dato da pregiudizi e stereotipi frutto di modelli culturali omologanti transitati nel senso comune delle classi dirigenti sarde e del ceto medio istruito) tra lingua-ghetto chiusa all'interno della cultura tradizionale agro-pastorale, per portarla invece a veicolare qualsiasi aspetto e contenuto che compone il caleidoscopico mosaico della cultura contemporanea dei cittadini della Sardegna, degli anni 2000. Non un rifiuto tout-court del mondo tradizionale o della cultura della memoria, ma certamente una svolta rispetto alla banalizzazione che una lingua minore debba per forza esprimere solo una civiltà subordinata, quasi fuori dal tempo, per ripiegarsi in usi, stilemi e conformismi che sono poi fossilizzati nella formula del "dialetto" incapace di esprimere i sentimenti, le aspirazioni e le complessità tipiche di una società che deve andare invece al passo con i tempi. In questo senso il nemico del

multilinguismo e il monolinguisma esasperato che è stato proposto per decenni all'interno delle istituzioni scolastiche italiane. Un programma di educazione monoglotta che è stato il più fiero avversario non solo dei dialetti o delle lingue minoritarie riconosciute, ma anche di un apprendimento sereno ed efficace delle lingue straniere.

Per questo, la strategia multilingue va incontro sia alla salvaguardia del sardo e delle lingue alloglotte, sia a un più capillare ed efficace uso delle lingue straniere di grande comunicazione europea e internazionale.

Avendo subito la lingua sarda tale processo di tipizzazione dialettale ed essendo stata in larga parte sclerotizzata quale "lingua della memoria", ma non utilizzata pienamente quale risorsa didattica veicolare, è prioritario in questo momento storico indirizzare gli investimenti a favore della lingua nella direzione di un uso privilegiato nel campo didattico della tecnologia e della comunicazione informatica al fine di favorire un'immagine più agile, snella, moderna, dinamica dello strumento comunicativo. In questo contesto è oltremodo importante l'implementazione della lingua che dovrà necessariamente osservare le norme linguistiche di riferimento a carattere sperimentale per la lingua sarda scritta dell'Amministrazione regionale, in quanto la mancanza di norme condivise, che siano lasciate all'arbitrio del personalismo dei singoli traduttori o scrittori, porterebbe a una percezione del codice linguistico individuale di ognuno come anarchico, personale, di poco prestigio e infine "dialettale".

Pertanto, l'Assessorato della Pubblica Istruzione, Beni Culturali, Spettacolo, Informazione e Sport, indica quale prioritari tre obiettivi da raggiungere attraverso la realizzazione del programma Scuola Digitale:

- 1) Sperimentazione della lingua sarda quale lingua veicolare dei contenuti didattici all'interno del Progetto Scuola Digitale in misura significativa e rilevante;
- 2) Inserimento di quote di argomenti di cultura sarda all'interno dell'offerta dei contenuti;
- 3) Realizzazione di contenuti, applicativi e programmi che utilizzino la lingua sarda in forma veicolare, in particolare nell'ambito didattico-tecnologico;

1) La sperimentazione del sardo quale lingua veicolare, accanto all'italiano e all'inglese, sarà realizzata nei seguenti lotti:

Lotto n°1 – Ambito di linguaggi per le scuole primarie e secondarie – 100 %

Lotto n°4 – Ambito storico-geografico – 100 %

Lotto n°5 – Ambito artistico musicale per tutti i gradi e gli ordini di scuole – 100 %

Lotto n°11 - Lingua e cultura sarda – 100 %

Lotto n°12 - Contenuti sperimentali relativi allo spazio di autonomia didattica riservato alla Regione – 100 %

L'implementazione delle lingue dovrà necessariamente osservare le norme linguistiche di riferimento a carattere sperimentale per la lingua sarda scritta dell'Amministrazione regionale.

2) Inserimento di quote di argomenti di cultura sarda all'interno dell'offerta dei contenuti sarà realizzata nei seguenti lotti:

Lotto n°1 – Ambito di linguaggi per le scuole primarie e secondarie - 20 %

Lotto n°4 – Ambito storico-geografico – 20%

Lotto n°5 – Ambito attività musicale – 20%

Lotto n°11 - Lingua e cultura sarda – 100 %

Lotto n°12 - Contenuti sperimentali relativi allo spazio di autonomia didattica riservato alla Regione – 100 %

3) Realizzazione di contenuti, applicativi e programmi che utilizzino la lingua sarda in forma veicolare, in particolare nell'ambito.

Progetti del Piano Triennale Lingua Sarda 2011-2013 da adattare ai contenuti della piattaforma di Scuola Digitale. L'implementazione delle lingua dovrà necessariamente osservare le norme linguistiche di riferimento a carattere sperimentale per la lingua sarda scritta dell'Amministrazione regionale.

Strumenti informatici per la lingua sarda

1) Manutenzione e aggiornamento del CROS (Correttore Regionale Ortografico Sardo)

Abstract

Le possibili linee di sviluppo del correttore automatico della lingua sarda sono:

- aumento dei termini riconosciuti;
- affinamento delle regole ortografiche e di generazione dei lemmi;
- miglioramento dei suggerimenti in caso di errori;
- aggiornamento alle nuove versioni dei sistemi operativi;
- supporto a nuovi programmi di videoscrittura;
- integrazione con altri strumenti linguistici (per esempio thesaurus, hyphenator, ecc., vedere 1.1, 1.2, 1.3; lo sviluppo di questi strumenti posso essere disgiunti);
- eventuale rimozione di bug o errori di programmazione.

Sinossi tecnica

Lo sviluppo del CROS proseguirà utilizzando le tecnologie già consolidate e testate durante le prime due versioni del correttore. In particolare si seguirà l'evoluzione e gli aggiornamenti resi disponibili dalla libreria Hunspell. Inoltre, tramite lo strumento di collaborazione e monitoraggio "Faeddos", si potrà vedere in tempo reale l'aggiornamento e la crescita del database linguistico di riferimento oltre che l'evoluzione delle regole di generazione dei lemmi.

Lo strumento di collaborazione "Faeddos" potrà essere ulteriormente affinato per dare la possibilità di aggiungere nuove informazioni ai vari lemmi (definizioni, fraseologia, errori di scrittura tipici, scrittura nelle grafie tradizionali del sardo, ecc.), per fornire statistiche aggiornate in tempo reale sul database linguistico, per permettere una comunicazione più rapida dentro il team di lavoro, ecc.

Particolare rilievo verrà dato all'aggiornamento del correttore per garantire la compatibilità con le nuove versioni dei programmi (Libre/OpenOffice), per la compatibilità con ulteriori programmi (video-scrittura, browser, ecc.), per il miglioramento sulle piattaforme web (versione ottimizzata per piattaforme mobile, maggiore compatibilità con i browser e con le nuove tecnologie e linguaggi web, adattamento per la piattaforma tecnologica regionale e integrazione con gli altri strumenti web dell'amministrazione).

1.1) *Hyphenation*

Abstract

Strumento per la divisione automatica delle parole in sillabe. Questo strumento è utile per la realizzazione di documenti tecnici o professionali (documenti amministrativi, libri, brochure, manifesti) in cui è necessario utilizzare la divisione delle parole e le "andate a capo" in maniera esatta. Può essere integrato in diversi programmi di videoscrittura come OpenOffice, LaTeX, ecc.

Sinossi tecnica

Lo strumento verrà sviluppato con tecnologia standard open source *libhyphen*, che è supportata dai principali applicativi di video-scrittura, impaginazione, sistemi editoriali, browser e sistemi web, sistemi mobile, sistemi operativi, ecc, oltre che essere integrabile con il Correttore Ortografico Sardo. Tale libreria permette di sviluppare regole di "andata a capo" personalizzate e adattate al 100% alla lingua considerata. La flessibilità consentita da questo strumento permette per esempio di creare delle regole in base alla lunghezza delle parole, al numero di sillabe, alla presenza di accenti, alla presenza di lettere doppie, alla presenza di grafemi composti (per esempio "tz", "gl", "ch", ecc.), alla presenza di segni tipografici o di punteggiatura particolari (per esempio "-", ".", ecc.).

L'utilizzo delle tecnologie a "codice aperto" consentiranno inoltre in futuro l'aggiornamento costante allo stato dell'arte della tecnologia, l'adattamento ad altre piattaforma o sistemi, la definizione di nuove regole, ecc.

1.2) *Thesaurus*

Abstract

Il Thesaurus è lo strumento che permette di implementare un dizionario di sinonimi e contrari. Esso permette di visualizzare in maniera automatica i sinonimi e i contrari di una parola selezionata. Può essere integrato in vari programmi di video-scrittura e sistemi operativi.

Sinossi tecnica

Il Thesaurus sarà sviluppato su tecnologia open source MyThes, che è compatibile con i più diffusi programmi di video-scrittura e sistemi operativi, oltre che integrabile con il Correttore Ortografico Sardo. Tale tecnologia permetterà di gestire i sinonimi e i contrari della lingua sarda e darà suggerimenti per i vari termini, classificati per parte grammaticale (nomi, aggettivi, verbi, ecc.). L'integrazione con il Correttore Ortografico permetterà inoltre di offrire i sinonimi e i contrari anche per le forme coniugate e derivate dei lemmi, indicando le corrispondenti forme coniugate e derivate.

Questo strumento risulta particolarmente utile sia nella redazione di documenti, report, racconti, e quant'altro, sia nella didattica della lingua, per permettere la conoscenza per ogni termine dei relativi sinonimi (compresi i geosinonimi) e i contrari. La realizzazione di tale strumento peraltro si integrerebbe con lo strumento di collaborazione "Faeddos" ed andrebbe ad aggiungere le informazioni sui sinonimi e contrari all'interno del database linguistico. Inoltre potrebbe essere un passo preparatorio per la realizzazione di un dizionario dei sinonimi e dei contrari della lingua sarda o per un dizionario generale avente anche la parte relativa ai sinonimi e contrari.

1.3) Correttore grammaticale

Abstract

Il correttore grammaticale permette di analizzare dei testi dal punto di vista grammaticale, evidenziando eventuali errori quali concordanza verbo/soggetto, posizione degli aggettivi nella frase, concordanza di genere e numero tra articoli, nomi e aggettivi, ecc. Può inoltre fornire suggerimenti di carattere stilistico, come la sostituzione di espressioni mutuata da altre lingue con espressioni o modi di dire tipici del sardo. Può essere integrato in vari programmi di video-scrittura e sistemi operativi.

Sinossi tecnica

Il correttore grammaticale sarà sviluppato con tecnologie open source e sarà integrato con il Correttore Ortografico Sardo. Il sistema sarà disponibile per i programmi di video-scrittura più

utilizzati per i vari sistemi operativi, nonché per il web. Il sistema potrà caricare molte diverse regole grammaticali o stilistiche, specifiche per la lingua sarda, in maniera da poter correggere gli errori più comuni.

Per la realizzazione si utilizzeranno i programmi e le librerie open source più adatta per le varie piattaforme e sistemi. Si svilupperanno le grammatiche, l'insieme delle regole e le varie osservazioni stilistiche.

Tale strumento risulta particolarmente utile sia per la redazione di testi, documenti formali, racconti, ecc. sia nell'utilizzo didattico in quanto permette l'identificazione e la correzione dei principali errori stilistici e grammaticali del sardo.

2) Dizionario completo della lingua sarda (con glossari specialistici, multi-piattaforma)

Abstract

Il dizionario completo della lingua sarda è uno strumento di consultazione di tutto il patrimonio della lingua sarda. Contiene definizioni, esempi d'uso dei termini, proverbi, modi di dire, e traduzioni nelle principali lingue europee. Potrà essere consultato tramite una applicazione per PC o telefoni cellulari di ultima generazione, o tramite sito web interattivo.

Sinossi tecnica

L'applicativo sarà sviluppato per i principali sistemi operativi (compresi quelli mobile) oltre che per il web. Il programma avrà tecnologie standard compatibili con la maggior parte dei sistemi informatici esistenti. Sostanzialmente il dizionario sarà un'interfaccia *user-friendly* del database linguistico precedentemente costituito, formato dai vari lemmi, definizioni, sinonimi, contrari, fraseologia, citazioni, proverbi, pronunce, traduzioni, eventuali grafie tradizionali alternative, ecc.

L'applicativo permetterà la ricerca dei termini o la consultazione libera, darà la possibilità all'utente di personalizzare la propria navigazione, ed eventualmente di aggiungere in maniera interattiva ulteriori proverbi, citazioni, fraseologia o pronunce.

Il dizionario potrà inoltre fornire le forme coniugate e derivate dei vari lemmi ed eventualmente le registrazioni delle varie pronunce.

3) Dizionario online delle altre varietà linguistiche (ver. semplificata rispetto al punto 2)

Abstract

Il dizionario delle altre varietà linguistiche presenti sul territorio sardo (tabarchino, sassarese, gallurese, algherese) è uno strumento necessario per tutelare e promuovere le lingue alloglotte della Sardegna, raccogliendo e normalizzando il patrimonio linguistico locale.

Sinossi tecnica

L'applicativo sarà sviluppato adattando quello per la lingua sarda (punto 2).

4) Correttore ortografico delle altre varietà linguistiche (ver. semplificata)

Abstract

Il correttore ortografico è lo strumento che permette di analizzare le parole di un testo e verificare la correttezza ortografica della stessa. In caso di parole non corrette, permette di visualizzare una lista di parole che sono simili alla parola errata. Il correttore ortografico per le varietà alloglotte è uno strumento per valorizzare e per normalizzare l'utilizzo secondo le regole del dizionario di cui sopra (la definizione del dizionario è necessariamente propedeutica alla realizzazione del correttore). Il correttore ortografico può essere integrato in programmi di videoscrittura (OpenOffice, Abiword, ecc.), browser (Mozilla Firefox, Google Chrome, ecc.), sistemi operativi (Linux, Mac OS X, ecc.) e siti web.

Sinossi tecnica

Lo sviluppo del Correttore Ortografico per le lingue alloglotte utilizzerà le tecnologie già consolidate e testate durante la realizzazione del Correttore Ortografico Sardo. In particolare si utilizzerà la tecnologia resa disponibile dalla libreria Hunspell. Inoltre, tramite lo strumento di collaborazione e monitoraggio "Faeddos", opportunamente adattato per le lingue alloglotte, si potrà vedere in tempo reale l'aggiornamento e la crescita del database linguistico di riferimento oltre che l'evoluzione delle regole di generazione dei lemmi.

5) Dizionario minimo sardo/italiano-italiano/sardo (multi-piattaforma)

Abstract



Una applicazione multi-piattaforma per la traduzione veloce dall'italiano al sardo e viceversa (o altre lingue europee) è uno strumento utile per l'utilizzo quotidiano e per l'apprendimento e l'impiego della lingua sarda. Grazie al fatto di essere multi-piattaforma potrà essere utilizzato in diversi ambiti sia nei normali PC che sui cellulari di ultima generazione.

Sinossi tecnica

L'applicativo fornirà un semplice dizionario bilingue in cui alla lista dei lemmi in sardo e in italiano verranno fatti corrispondere i relativi termini nell'altra lingua. Il programma sarà compatibile con la maggior parte dei sistemi operativi sia per PC che per il settore mobile, e sarà sviluppato con tecnologie standard.

In particolare si terrà conto della facilità di utilizzo e della possibilità di effettuare ricerche in entrambe le lingue e il poter aggiungere informazioni grammaticali di base.

Tale strumento si ritiene particolarmente utile sia per il parlante esperto che può avere dubbi sulla grafia standard di una parola, sia per i neoparlanti che hanno bisogno di incrementare il loro vocabolario personale. Dunque tale strumento appare di eccezionale valore sia per la normalizzazione, sia per la didattica del sardo.

6) Sistemi di digitazione predittivi (T9 o similari)

Abstract

Il "T9" è un sistema di digitazione veloce per tastiere che non possiedono tutti i caratteri (come per esempio cellulari, telefoni, palmari, telecomandi, ecc.). Esso permette una maggiore velocità di scrittura ottimizzando il numero di digitazioni per produrre un testo. Da tener presente che i telefonini di nuova generazione (ad esempio basati sul S.O. Android di Google), i cosiddetti *smartphones*, permettono di superare questa tecnologia con l'inserimento dei test attraverso delle tastiere virtuali.

Sinossi tecnica

Il programma sarà sviluppato con tecnologie standard compatibili con i più comuni sistemi operativi per dispositivi *mobile*. In particolare si svilupperà la lista dei termini riconosciuti dal sistema e in contemporanea una lista di frequenza dei vari termini, generata a partire da un corpus linguistico formato da pubblicazioni di vario genere.



Unione europea
Fondo sociale europeo



Repubblica Italiana



REGIONE AUTONOMA DELLA SARDEGNA



obiettivo competitività regionale e occupazione

Il sistema proporrà il completamento della digitazione della parola con le parole presenti nella base di dati, utilizzando l'ordine fornito dalla lista di frequenza. Il programma sarà personalizzabile dall'utente su vari aspetti e l'intera interfaccia utente dell'applicazione sarà in sardo. L'utente potrà inoltre inserire dei termini in un proprio vocabolario personale per poter ulteriormente estendere il numero di parole riconosciute dal sistema.

La realizzazione di un sistema di digitazione deve essere preceduto da alcune che sono di natura prettamente linguistica. Si inizia con l'analisi statistica su del materiale linguistico consistente (*corpus*), selezionato e normalizzato. Il corpus va esaminato attentamente per comprendere il comportamento degli elementi linguistici che serviranno per comporre le regole di predizione (frequenza dei caratteri, digrammi, trigrammi, parole e frasi). Un'analisi lessicale su un corpus abbastanza vasto non è mai stata eseguita per la lingua sarda (tale studio può avere altre ricadute molto interessanti, tipo: la creazione di un dizionario/vocabolario tascabile; anticipare i lavori necessari per la creazione delle librerie linguistiche; fornire informazioni per la creazione di corsi di lingua sarda; ecc.; vedere il punto 8 del presente documento). Segue la creazione di una serie di regole/algoritmi di componimento dei testi in lingua sarda. Le regole dovranno mirare alla creazione di un sistema predittivo che minimizzi il tempo e lo sforzo necessario per l'inserimento dei testi attraverso una tastiera da telefono cellulare e/o tastiera virtuale. Dovranno essere prese in considerazione diverse metodologie e tecniche di ottimizzazione dei movimenti sulla tastiera e di predizione del linguaggio. Il database di base dei lemmi, insieme alle regole e algoritmi opportunamente codificati, costituiranno il sistema informatico che andrà inserito nel cellulare. Infine vanno eseguite le prove e il collaudo del sistema che va testato da un campione il più possibile rappresentativo dell'utenza reale. Il testing del sistema andrà eseguito in più fasi che potranno essere, in parte, effettuate in parallelo a quelle di sviluppo.

7) Traduttore automatico

Abstract

Il traduttore automatico permette di tradurre dal sardo ad altre lingue europee e viceversa, testi (anche complessi), rispettando i costrutti linguistici e grammaticali della lingua in cui viene tradotto il testo. Il traduttore, tramite un'analisi grammaticale e logica del testo di partenza, permette di tradurre il senso del testo, evitando dunque la traduzione parola per parola tipica di strumenti meno avanzati. Tale sistema può essere implementato sia come programma per PC sia come sito Internet.

Sinossi tecnica

La traduzione basata su regole, che comprende la traduzione basata sul principio del trasferimento (transfer) e quella basata sull'interlingua (lingua ausiliaria internazionale), è al giorno d'oggi il paradigma prevalente per i sistemi di traduzione automatica. I più famosi sistemi di traduzione automatica attuali sono tutti basati su regole. Fra questi i più conosciuti sono sicuramente Google e Babelfish (usato da Altavista).

Il processo traduttivo suddiviso è generalmente in tre fasi. Nella prima fase (nota come fase di *analisi*), il sistema esegue il *parsing* delle frasi del testo di partenza e le trasforma in diagrammi ad albero (morfologici, sintattici e/o semantici). Nella seconda fase (detta di *trasferimento*), gli alberi sintattici creati per il testo di partenza vengono trasformati in altrettanti alberi con la struttura sintattica della lingua d'arrivo. Nella terza fase (chiamata di *generazione* o sintesi), le parole della lingua di partenza vengono tradotte nella lingua finale e inserite nell'albero d'arrivo seguendo le regole sintattiche proprie della lingua medesima al fine di creare frasi di senso compiuto.

Una tecnica particolare della traduzione automatica basata su regole è quella che utilizza un linguaggio intermedio (interlingua). Rispetto al sistema di traduzione diretta, parola per parola, in questo caso la lingua d'origine del testo da tradurre sarebbe trasferita a una lingua intermedia, la cui struttura è indipendente da quella della lingua originale e da quella della lingua finale. Il testo nella lingua di destinazione è ottenuto a partire dalla rappresentazione del testo nella lingua intermedia.

Il principale punto debole di questa tecnologia è che la quantità di regole su cui si basa il sistema è naturalmente limitata. Per ottenere risultati migliori, gli autori del testo da tradurre dovrebbero adeguare il loro stile di scrittura, circostanza assai difficile da proporre.

Oggi gli applicativi più complessi offrono la possibilità di fissare parametri specifici, secondo il settore in cui si esegue la traduzione, per ottenere risultati più accurati. Ad esempio si può limitare la quantità di sostituzioni permesse oppure selezionare i nomi propri e aiutare il software a riconoscere la terminologia e la fraseologia più ricorrenti. Queste tecniche sono particolarmente utili in quei settori dove si impiega un linguaggio formale basato su moduli, come i documenti legali e amministrativi o le previsioni del tempo e gli annunci nelle stazioni ferroviarie e negli aeroporti.

In questo ambito può essere utile vedere le esperienze maturate dai progetti "Stilven" dell'Università Ca' Foscari di Venezia (traduzione dalla lingua veneta alla lingua inglese e viceversa) e "Jude" (traduzione dalla lingua italiana a quella friuliana) realizzato dalla società *Serling Soc. Coop. Servizis pe lenghe furlane*.

8) Database linguistico

Abstract

L'archivio linguistico è una base dati che permette di archiviare e gestire in maniera complessa il patrimonio linguistico della lingua sarda, permettendo di classificare e ricercare informazioni da una serie di fonti eterogenee (dizionari, testi letterari, testi amministrativi, ecc.). Il database può essere impiegato come base per eventuali ricerche scientifiche, come riferimento per la realizzazione di altri strumenti, e come base dati aperta (eventualmente accessibile anche tramite siti internet) per lo studio della lingua sarda.

Sinossi tecnica

Il database linguistico è un programma che permette la catalogazione e la gestione dei corpus linguistici di una lingua.

Riportiamo qui di seguito la definizione tratta da Wikipedia: «La linguistica dei *corpora* rappresenta una disciplina scientifica matura, che in cinquant'anni è riuscita a conquistare una posizione di indiscussa centralità nel panorama scientifico nazionale e internazionale, e in cui sono attivi ormai innumerevoli gruppi di ricerca. Servendosi di strumenti di analisi quantitativa e statistica, esplora le regolarità linguistiche che emergono dai testi e che costituiscono la base per la descrizione della struttura del linguaggio. L'obiettivo principale è quello di sviluppare modelli del funzionamento del linguaggio umano che possono essere tradotti in programmi eseguibili dal calcolatore e che gli consentono di acquisire le competenze necessarie per comunicare direttamente nella nostra lingua. Si tratta di un ambito di ricerca e di lavoro tipicamente interdisciplinare. Infatti, da un lato si intreccia con l'informatica e con l'ingegneria, per qualificare le attività rivolte allo sviluppo di applicazioni informatiche basate sulle tecnologie del linguaggio, mentre dall'altro lato essa dialoga, appunto, con la linguistica, le scienze cognitive, la psicologia, la filosofia e le scienze umane in generale, con le quali condivide l'obiettivo di indagare la struttura, il funzionamento e l'uso del linguaggio e il suo rapporto con le altre facoltà cognitive dell'uomo.» visibile digitando la voce 'corpus' dalla home page della versione in italiano.

Tecnicamente parlando l'archivio linguistico dovrà essere sviluppato con tecnologie standard e sarà compatibile con i più diffusi sistemi operativi e browser. Avrà un'interfaccia web che permetterà di visualizzare, compiere ricerche, trovare le relazioni, ecc. all'interno del corpus linguistico di riferimento.

La raccolta del materiale può essere fatto da elementi esistenti come pagine web, e-mail, libri, riviste, registrazioni ecc. Il corpus deve garantire una rappresentatività della lingua in termini di *dimensioni* (numero elevato di testi), *autenticità* (deve rappresentare la reale produzione linguistica) e *bilanciamento* (deve avere una caratterizzazione adeguata come origine e tipologia). Data la natura del materiale esistente il problema maggiore potrebbe essere l'omogeneizzazione dei testi per facilitarne il successivo trattamento (la dimensione dovrebbe essere sufficiente in quanto si riesce a raccogliere agevolmente testi moderni e contemporanei per oltre 15 milioni di parole).

Il database potrà avere vari livelli di fruizione per cui, in base a delle credenziali di accesso, sarà possibile accedere a differenti risorse linguistiche. Potrà essere prevista eventualmente anche una parte pubblica aperta a tutti. Il database potrà comprendere inoltre contenuti multimediali, quali filmati, registrazioni audio, immagini, collegamenti, ecc. catalogati e collegati ai corrispondenti dati testuali.

Tale strumento appare di importanza primaria per lo studio a livello accademico della lingua sarda, per lo sviluppo e il perfezionamento di tutti gli altri strumenti informatici e cartacei (dizionari, antologie, raccolte di fraseologia, ecc.), per la didattica nelle scuole di ogni ordine e grado, fornendo in maniera catalogata e ordinata delle informazioni e dei materiali utili alla realizzazione di lezioni, approfondimenti, letture, ascolti, ecc.

Infine tale opera dovrà essere costantemente aggiornata nel tempo con l'inserimento costante di nuovi materiali riguardanti la lingua sarda che l'Amministrazione Regionale acquisirà.

9) Text-To-Speech (TTS)

Abstract

Gli strumenti di *Text-to-speech* sono dei programmi che permettono di riprodurre con una voce sintetizzata dei testi scritti. Tali programmi, detti anche *language libraries*, sono fondamentali per aumentare l'accessibilità dell'utilizzo del PC da parte di categorie svantaggiate e per l'apprendimento della pronuncia. Inoltre permettono l'ascolto di e-book e testi elettronici (email, documenti di testo, ecc.). Tali sistemi possono essere integrati in programmi per PC (giochi multimediali interattivi, dizionari, corsi di lingua), dispositivi portatili (cellulari, e-book reader, ecc.), centralini telefonici, annunciatori per luoghi pubblici, avatar parlanti, siti web (soprattutto per l'accessibilità, v. Legge Stanca) e via dicendo. Lo sviluppo di questi strumenti deve essere proceduto da un'analisi sistematica e mirata della fonetica della lingua sarda che porterà ad ottenere un archivio linguistico

elettronico composito ed eterogeneo, aperto a sviluppi e integrazioni futuri. Gli obiettivi principali sono l'individuazione degli elementi fonetici e fonologici della lingua sarda, la loro registrazione ed elaborazione, l'inserimento in un sistema informatico complesso e la predisposizione di regole che ne permetta l'utilizzo attraverso programmi linguistici specifici e mirati. In particolare il lavoro è propedeutico alla realizzazione di un sistema ancora più avanzato di riconoscimento vocale (*Automatic Speech Recognition* => ASR) che permette di convertire una registrazione o un dettato in testo (*Speech-to-text*).

Sinossi tecnica

Il progetto mira a realizzare un prototipo elettronico che possa riprodurre il parlato partendo da testi scritti (sintesi vocale). Per *sintesi vocale* si intende la tecnica utilizzata per riprodurre artificialmente la [voce umana](#). Il lavoro è volto a creare un corpus di dati fonetici registrati, regole linguistiche e archivi informatici di base mirati, tra le altre cose, allo sviluppo futuro di un software di sintesi vocale per la lingua sarda.

La ricerca di base per la creazione delle biblioteche linguistiche consisterà soprattutto in un'analisi sistematica e mirata della fonetica della lingua sarda che porterà ad ottenere un archivio linguistico elettronico composito ed eterogeneo, aperto a sviluppi e integrazioni futuri.

Il fatto di avviare una ricerca di raccolta dati che costituiscano gli elementi di base per un sistema di sintesi vocale, dal momento che in pratica la sintesi vocale può essere realizzata attraverso la concatenazione di registrazioni di parti vocali memorizzate in un [database](#) ([fonemi](#) singoli o sequenze di fonemi, parole o intere frasi), significa costituire un database linguistico provvisto di informazioni sulla fonetica, sulla fonologia e su alcuni elementi di prosodia.

Il progetto è caratterizzato da una forte interdisciplinarietà tra la linguistica e l'informatica. In un lavoro di ricerca come quello necessario per la realizzazione di un sistema di sintesi vocale, le due discipline operano in costante interazione: la linguistica lavora alla costituzione di un corpus di dati fonologici e l'informatica cura la parte di registrazione ed elaborazione dei suoni, la catalogazione sistematica attraverso un database dedicato e l'applicazione delle informazioni al sistema di trasformazione dal testo scritto alla voce artificiale.

I campi di applicazione della sintesi vocale sono vari e di grande interesse. Molte università importanti nel mondo sono dotate di laboratori linguistici che si occupano di questa disciplina e di quella, ad essa collegata, di riconoscimento vocale. Alcuni siti web dedicati alla diffusione di notizie e in generale dell'informazione utilizzano la sintesi vocale per convertire i bollettini informativi scritti in modalità audio per una fruizione di tipo mobile. Le tecniche di sintesi vocale trovano impiego nell'industria del tempo libero, in particolare nella produzione di videogiochi e simili. Non ultimo, tali

tecniche costituiscono uno strumento di assistenza tecnologica di grande importanza che può essere applicato al fine di eliminare barriere ambientali in un'ampia gamma di situazioni di esclusione sociale (persone affette da problemi alla vista, [dislessia](#) o da altri problemi di lettura, persone con gravi problemi vocali o di fonazione).

10) Traduzione di programmi *open source*

Abstract

La traduzione di applicativi di largo uso (OpenOffice, LibreOffice, Linux Ubuntu, Mozilla Firefox, ecc.) in lingua sarda è uno tra i modi più incisivi per promuovere l'uso della lingua sarda, in quanto ne incoraggia l'impiego in tutti quei contesti in cui vengono utilizzate le tecnologie informatiche. Un aspetto non secondario è la creazione di un glossario specialistico di termini informatici. I programmi che si possono tradurre vanno da quelli di video-scrittura (OpenOffice/ LibreOffice) a quelli di navigazione in Internet (Firefox, Chrome), da programmi didattici ed educativi a quelli ludici (videogiochi), fino ad interi sistemi operativi.

Sinossi tecnica

La traduzione dei testi descrittivi di un applicativo open source non presenta particolari difficoltà dal punto di vista tecnico. Molti programmi hanno già un sistema predisposto per la traduzione dei campi testuali. Le difficoltà sono soprattutto di natura linguistica poiché trattasi di glossari specialistici inesistenti per quanto riguarda la lingua sarda (esistono alcuni tentativi "fai da te" anche lodevoli, ma niente di ufficiale). L'altra difficoltà è di natura organizzativa dovuta alla quantità ingente di testi da tradurre, la natura specialistica degli stessi e il continuo aggiornamento che deve essere fatto per non rendere obsoleto il lavoro svolto in precedenza.

11) Opere didattiche di apprendimento della lingua sarda

Opere didattiche in lingua sarda per l'apprendimento della lingua sarda nelle scuole di ogni ordine e grado, con adattamento del multimediale all'impiego in rete, nel rispetto dei requisiti previsti dalle norme vigenti sull'accessibilità ai servizi informatici e telematici della Pubblica Amministrazione;



12) Realizzazione grafica e video audio di Topolino in lingua sarda

Sorigheddu - Topolino in lingua sarda

La politica linguistica regionale promuove inoltre l'apprendimento e l'insegnamento della lingua sarda all'interno di un percorso educativo plurilingue. Il suddetto percorso dovrà costituire parte integrante della formazione a una cittadinanza europea attiva, e al consolidamento di una specificità storico - identitaria della comunità regionale. La Regione, riconoscendo sensibile tale argomento, ritiene di dover adottare un metodo graduale di diffusione dell'insegnamento della lingua.

Rientrano dunque in quest'ottica tutti quei progetti atti a sensibilizzare i più giovani all'uso della lingua, a coinvolgerli e stimolarli, facendo leva sul principio di riconoscimento del proprio popolo.

In questo senso si inserisce la sinergia tra la lingua sarda e l'innovazione. Innovazione di contenuti, di metodologia e sostanza. A questo scopo la Regione incoraggia progetti-pilota nelle autonomie scolastiche interessate a sperimentare l'insegnamento in orario curricolare della lingua sarda, e va di pari passo l'interesse nel creare nuovi strumenti didattici, in linea con l'idea di base e gli obiettivi preposti.

In questo contesto si inserisce il progetto editoriale che vede la traduzione in lingua sarda della nota pubblicazione e dei cartoni della Walt Disney, **Topolino**.

Il personaggio che ha accompagnato l'infanzia di più generazioni in tutto il mondo, potrebbe ora diventare una realtà nostrana, un veicolo della nostra cultura e della nostra lingua. Potenziale strumento didattico per i più piccoli, eccellente oggetto di culto per i più grandi.

La lingua sarda verrebbe valorizzata e diffusa in un modo del tutto inaspettato, avvalendosi della popolarità di **Topolino**.

L'iniziativa sarà realizzata con la Disney o con i concessionari del marchio e dei diritti in Sardegna.