



# Integrazione di tecnologie per la traduzione e CLIL nella creazione di risorse didattiche per il sardo

GIANFRANCO FRONTEDDU<sup>1</sup>, IGOR DEIANA<sup>2</sup>

<sup>1</sup>UNIVERSITÀ DEGLI STUDI DI CAGLIARI [GIANFRANCO.FRONTEDDU@UNICA.IT](mailto:GIANFRANCO.FRONTEDDU@UNICA.IT)

<sup>2</sup>UNIVERSITÀ DEGLI STUDI DI CAGLIARI [IGOR.DEIANA@UNICA.IT](mailto:IGOR.DEIANA@UNICA.IT)<sup>1</sup>

## ABSTRACT (ITALIANO)

Il contributo propone una riflessione metodologica sull'uso integrato delle tecnologie per la traduzione nella produzione di materiali didattici per l'insegnamento del sardo in ambito scolastico, con particolare riferimento alla metodologia CLIL (*Content and Language Integrated Learning*). A partire dal quadro normativo nazionale e regionale che promuove l'inserimento del sardo in orario curricolare (L. 482/1999; L.R. 22/2018; avvisi INSULAS e FRAILES), il lavoro affronta la problematica della scarsità di risorse didattiche aggiornate per una lingua a risorse limitate. Dopo una sintesi dei principali approcci alla traduzione automatica (basata su regole, su corpus e neurale), il contributo analizza il ruolo del post-editing come pratica di mediazione linguistica in linea con il QCER.

In questo contesto, viene discusso l'uso del sistema di traduzione automatica *rule-based* Apertium italiano-sardo, integrato con strumenti di traduzione assistita *open source*, in particolare OmegaT, per la creazione di Unità di Apprendimento CLIL. La proposta metodologica è illustrata attraverso un caso di studio sperimentato nell'ambito del Master di II livello per la Didattica della Lingua Sarda dell'Università di Cagliari, relativo a un'UdA di storia per la scuola secondaria di primo grado. L'esperienza mostra come la combinazione tra traduzione automatica e TAO consenta di aumentare significativamente velocità ed efficienza nella produzione dei materiali, favorendo il riuso, l'adattamento e la standardizzazione di risorse digitali in lingua sarda. Se adottata su larga scala, la metodologia permette inoltre la condivisione delle memorie TMX generate, utili sia alla coerenza terminologica sia allo sviluppo dell'ecosistema digitale del sardo, anche in prospettiva di addestramento di sistemi IA e motori NMT.

**Parole chiave:** Traduzione automatica; Traduzione assistita (TAO); Post-editing; CLIL; Didattica del sardo

## ABSTRACT (ENGLISH)

*Integrating translation technologies and CLIL in the creation of teaching resources for Sardinian.*

This paper offers a methodological reflection on the integrated use of translation technologies in the production of teaching materials for Sardinian in schools, with particular reference to the CLIL (Content and Language Integrated Learning) methodology. Drawing on the national and regional regulatory framework promoting the inclusion of Sardinian in the curriculum (Law 482/1999; Regional Law 22/2018; INSULAS and FRAILES notices), the paper addresses the issue of the scarcity of up-to-date teaching resources for a language with limited resources. After a summary of the main approaches to machine translation (rule-based, corpus-based, and neural), the paper analyzes the role of post-editing as a linguistic mediation practice in line with the Common European Framework of Reference for Languages (CEFR). In this context, the use of the rule-based machine translation system Apertium Italian-Sardinian, integrated with open-source computer-assisted translation tools, particularly OmegaT, for the creation of CLIL Learning Units is discussed. The proposed method is illustrated through a case study developed as part of the Master's Degree in Teaching the Sardinian Language at the University of Cagliari, involving a history course for middle school students. This experience demonstrates how the combination of machine translation and TAO significantly increases the speed and efficiency of material production, promoting the reuse, adaptation, and standardization of digital resources in the Sardinian language. If adopted on a large scale, the methodology also allows for the sharing of generated TMX memories, which are useful both for terminological consistency and for the development of the Sardinian digital ecosystem, including for training AI systems and NMT engines.

---

<sup>1</sup> L'articolo è il frutto di un lavoro congiunto e di una stretta collaborazione degli autori. All'interno di una concezione ed elaborazione comune e condivisa in tutti i suoi aspetti, sono tuttavia da attribuire a Gianfranco Fronteddu i paragrafi 3, 4 e 6 e a Igor Deiana il paragrafo 2. I paragrafi 1, 5 e 7 sono da attribuire a entrambi gli autori.

**Keywords:** Machine translation; Computer-assisted translation (CAT); Post-editing; CLIL; Teaching Sardinian

## 1. INTRODUZIONE

Il contributo propone una riflessione sull'uso integrato della traduzione automatica e della traduzione assistita per la produzione di materiali didattici finalizzati all'uso veicolare della lingua sarda per lo studio di contenuti disciplinari secondo la metodologia CLIL (*Content and Language Integrated Learning*). Muovendo dal quadro normativo e dalle politiche linguistiche regionali, il lavoro analizza il potenziale didattico di strumenti open source per lingue minoritarie. In particolare, viene presentata una metodologia applicativa che combina traduzione automatica basata su regole, *post-editing* e *CAT tools*. L'approccio è illustrato attraverso un caso di studio ideato e presentato in occasione delle attività del laboratorio "Laboratorio di didattica delle lingue e delle nuove tecnologie e metodologia CLIL", tenuto dagli autori del contributo, nell'ambito del Master di II livello per la Didattica della Lingua Sarda dell'Università degli Studi di Cagliari.

## 2. PRODUZIONE DI MATERIALE DIDATTICO PER LA DIDATTICA DEL SARDO IN CLIL

L'inserimento dell'insegnamento del sardo in orario curricolare in ambito scolastico trae fondamento giuridico sia dalla normativa statale a tutela delle minoranze linguistiche (L. 482/1999)<sup>2</sup> che dalla politica linguistica regionale (L.r. 22/2018)<sup>3</sup>(Marra 2021). Le iniziative proposte dalla Regione Autonoma della Sardegna (d'ora in avanti RAS) per il potenziamento dell'offerta formativa sono articolate nelle linee INSULAS ("INsegnamento Unico Lingue A Scuola") e FRAILES ("Fucine di Lingue sarde") che finanziano, rispettivamente, progetti per l'insegnamento e l'utilizzo veicolare del sardo in orario curricolare e laboratori didattici extracurricolari nelle scuole dell'infanzia, primaria e secondaria di primo e secondo grado (RAS 2025). Se già negli ultimi decenni le scuole dell'Isola hanno visto lo sviluppo di proposte didattiche per la lingua sarda, le iniziative più recenti hanno contribuito ad arricchire il panorama dell'offerta (Mereu, 2024). Sono un valido esempio alcune attività dei corsi INSULAS che si basano sulla metodologia CLIL, attraverso la quale l'apprendimento di una lingua straniera o minoritaria avviene in modo più efficace quando la stessa è veicolare di contenuti disciplinari; infatti, in questo modo si favorisce l'integrazione tra competenze linguistiche, metalinguistiche e conoscitive (Coyle et al., 2010). In questo quadro, l'uso della lingua sarda come lingua veicolare consente di collegare l'apprendimento formale delle strutture e del lessico con l'analisi di contenuti culturali e disciplinari, dando luogo a un percorso didattico che supera la semplice acquisizione di regole e che favorisce la costruzione di conoscenze integrate.

L'introduzione di questi percorsi rende evidente la necessità di disporre di materiali didattici per il sardo aggiornati e adeguati alla moderna didattica delle lingue, dei quali, purtroppo, al momento, vi è poca disponibilità (Micali 2023, Grosso et al. 2024). Come ricorda infatti Mereu (2024): "I materiali didattici a disposizione per l'insegnamento del sardo (e in sardo) sono ancora molto limitati e, in genere, si ricorre a dei testi già esistenti, adattati alle esigenze scolastiche". A questa criticità si aggiunge il fatto che la creazione del materiale necessario allo svolgimento delle attività richiede, oltre a specifiche competenze glottodidattiche, un notevole dispendio di tempo anche quando si trae spunto da testi autentici o ricavati da libri di testo in adozione.

Alla luce del quadro appena descritto, come sarà illustrato nei paragrafi 3 e 4, l'uso combinato della Traduzione Automatica in combinazione con la Traduzione Assistita dal Computer, insieme alle risorse offerte dalla didattica delle lingue (sillabi e referenziali), possono essere impiegati come supporti alla generazione di risorse per l'insegnamento curricolare del sardo.

---

<sup>2</sup> Stato Italiano. (1999). Legge 15 dicembre 1999, n. 482 – Norme in materia di tutela delle minoranze linguistiche storiche (G.U. n. 297, 20 dicembre 1999). Url: <https://www.mim.gov.it/lingue-di-minoranza> (cons. 31/12/2025)

<sup>3</sup> Regione Autonoma della Sardegna. (2018). Legge regionale 3 luglio 2018, n. 22 — Disciplina della politica linguistica regionale. Url: <https://leggiregionali.regione.sardegna.it/legge-regionale?data=3-7-2018&numero=22> (cons. 31/12/2025)

### 3. TRADUZIONE AUTOMATICA PER LA LINGUA SARDA: APERTIUM

La *traduzione automatica (d'ora in avanti TA)* o in inglese *Machine Translation (MT)* è l'espressione più diffusa per indicare i sistemi informatici che traducono da una lingua a un'altra, con o senza intervento umano, precedente o successivo al processo di traduzione (Monti, 2023).

Per classificare le principali tipologie di TA si distinguono i sistemi basati su regole (*rule-based machine translation, d'ora in poi RBMT*), in cui esperti umani definiscono esplicitamente risorse linguistiche e regole di trasferimento, e i sistemi basati su corpus (*statistical machine translation, d'ora in poi SMT*), il cui funzionamento dipende dall'addestramento su grandi quantità di dati paralleli utilizzati per la costruzione di modelli di traduzione e modelli linguistici (Martín-Mor et al., 2016). La traduzione automatica neurale (*neural machine translation, d'ora in avanti NMT*) rappresenta un'evoluzione di tali approcci ed è generalmente considerata lo stato dell'arte della TA (Castilho et al., 2017). Basata su reti neurali artificiali e modelli di *deep learning*, la NMT è in grado di elaborare testi completi apprendendo rappresentazioni linguistiche astratte, analogamente ai grandi modelli linguistici (*Large Language Models; d'ora in avanti LLM*) (Shormani, 2024; Wu et al., 2016). L'impiego di architetture *transformer* ha migliorato la fluidità e la coerenza testuale delle traduzioni, grazie alla capacità di catturare *pattern* di equivalenza tra lingue attraverso vasti corpora empirici (Elhamayed & Nour, 2025). Tuttavia, nonostante i progressi dell'Intelligenza Artificiale, d'ora in poi IA, la TA non ha ancora raggiunto un livello pienamente equivalente alla traduzione umana, soprattutto su testi estesi e a livello documentale (Mattioda et al., 2023).

La distinzione tra RBMT, SMT e NMT e la conoscenza delle rispettive caratteristiche computazionali, consente di comprenderne limiti e potenzialità e di adottare strategie d'uso adeguate (Haddow et al., 2022; Lakew et al., 2018). Studi comparativi mostrano, ad esempio, che i sistemi RBMT tendono a garantire una maggiore correttezza morfosintattica, ma risultano rigidi sul piano stilistico, mentre i sistemi NMT producono *output* più fluidi ma con errori lessicali specifici (Koponen et al., 2019).

In questo quadro, il *post-editing (Machine Translation Post-Editing, MTPE)* assume un ruolo centrale, inteso come revisione dell'*output* della TA (Balling et al., 2014). Si distinguono un *post-editing leggero*, volto alla correzione degli errori che ostacolano la comprensione, e un *post-editing completo*, che mira a un livello qualitativo comparabile a quello della traduzione umana. Si ricorda che invece, a monte, il *pre-editing* interviene sul testo di partenza tramite linguaggi controllati per prevenire errori (Riediger & Galati, 2023).

L'uso integrato di TA e *post-editing* consente una significativa riduzione di tempi e costi rispetto alla traduzione umana tradizionale. Studi e report settoriali indicano risparmi economici fino al 47% e riduzioni dei tempi di consegna fino al 50-60%, con incrementi di produttività compresi tra +38% e +66% (Kapusta et al., 2021). Tali vantaggi risultano particolarmente rilevanti nel contesto delle lingue minoritarie, dove la generazione rapida di contenuti digitali rappresenta una priorità strategica e metodologica (Terribile, 2024).

Nell'ambito della lingua sarda, un sistema di riferimento è Apertium, piattaforma di traduzione automatica basata su regole di trasferimento superficiale, particolarmente adatta alla traduzione tra lingue romanze (Tyers et al., 2017; Fronteddu et al., 2017). Il sistema è costituito da un motore centrale e da risorse linguistiche codificate in XML, tra cui dizionari morfologici monolingui e bilingui e file di regole, organizzate secondo una struttura modulare che riflette le tre fasi tipiche dei traduttori indiretti: analisi, trasferimento e generazione (Tyers et al., 2017; Fronteddu et al., 2017). Attualmente, per il sardo sono disponibili le coppie italiano-sardo e catalano-sardo, in entrambe le direzioni di traduzione (Tyers et al., 2017; Fronteddu et al., 2017). Le caratteristiche tecniche di Apertium hanno reso necessaria l'adozione di una norma ortografica coerente e sufficientemente attestata, individuata nella *Limba Sarda Comuna (LSC)*,<sup>4</sup> funzionale alla costruzione di risorse digitali e di un corpus adeguato (Fronteddu et al., 2017; Tyers et al., 2017). In termini di qualità, Apertium italiano-sardo presenta un tasso di errore pari al 9,9% su testi di carattere generale (Tyers et al., 2017).

Negli ultimi anni, l'ecosistema delle tecnologie per la traduzione automatica del sardo si è ampliato includendo servizi basati su modelli neurali, come MinT (Pombal et al., 2025), e progetti di ricerca quali LIMBA (Carta et al., 2024) e A BERT (Carta et al., 2025), orientati allo sviluppo di risorse digitali e strumenti di apprendimento basati su IA.

---

<sup>4</sup> Limba Sarda Comuna (LSC). Url: [https://www.regione.sardegna.it/documenti/1\\_72\\_20060418160308.pdf](https://www.regione.sardegna.it/documenti/1_72_20060418160308.pdf) (cons. 05/02/2026)

In un contesto didattico come nel caso oggetto del presente studio, Apertium italiano-sardo risulta particolarmente adatto poiché, grazie all'integrazione con CAT tools come OmegaT (vedi paragrafo 4), consente un *post-editing* più controllabile e funzionale alla produzione efficiente di materiali didattici di base.

#### 4. TRADUZIONE ASSISTITA (*COMPUTER-ASSISTED TRANSLATION*)

L'integrazione di strumenti di TAO risulta complementare all'uso della TA. La definizione di Traduzione Assistita dal Computer (*Computer Assisted Translation*) si riferisce all'utilizzo di programmi informatici a supporto della figura del traduttore nel processo di traduzione (Alsaawi, 2025).

Un caso di riferimento è quello di OmegaT,<sup>5</sup> che attualmente è uno dei programmi *open-source*<sup>6</sup> per la TAO, o Cat Tools, predisposti per la gestione di progetti per il sardo. Secondo quanto riportato nella guida aggiornata al 2020,<sup>7</sup> OmegaT è un software open source e gratuito sviluppato in Java, pensato per supportare traduttori professionisti e non professionisti nella gestione di progetti di traduzione tramite memorie di traduzione, glossari e corrispondenze parziali (*fuzzy match*) tra segmenti testuali. OmegaT non traduce autonomamente, ma offre un ambiente in cui il traduttore può ri-utilizzare e organizzare risorse linguistiche per migliorare la coerenza e l'efficienza del processo traduttivo. Dal punto di vista funzionale, OmegaT supporta la gestione di memorie di traduzione in formati standard come TMX (*Translation Memory eXchange*) e XLIFF (*XML Localization Interchange File Format*), l'uso di glossari con riconoscimento delle forme flesse, ed è integrabile con *plugin* (ad esempio, per collegarsi a motori di TA quali Apertium), oltre a fornire strumenti di allineamento, facilitando l'uso combinato in ambito didattico tra TAO e MT. Per il contesto del sardo, il supporto di Unicode<sup>8</sup> rende OmegaT particolarmente adatto, perché consente la configurazione manuale di coppie linguistiche e codici lingua anche non standardizzati, senza problemi con l'ortografia sarda.

#### 5. METODOLOGIA

In occasione delle attività svolte durante il "Laboratorio di didattica delle lingue e delle nuove tecnologie e metodologia CLIL" del Master di II livello in Didattica della Lingua Sarda organizzato nell'a.a 2024-2025 dal Dipartimento di Lettere, Lingue e Beni Culturali dell'Università degli Studi di Cagliari, con il co-finanziamento della RAS, è stata ideata e presentata una metodologia volta all'uso di strumenti digitali attualmente disponibili per la creazione di Unità di Apprendimento (d'ora in avanti UdA) da destinare all'insegnamento della lingua sarda e, nello specifico, al suo uso veicolare in modalità CLIL.

La metodologia in questione prevede la creazione di un'UdA, in linea con i descrittori del Quadro Comune Europeo di Riferimento per la conoscenza delle lingue, d'ora in avanti QCER, (Consiglio d'Europa 2002, 2020) e con i principi della didattica per competenze (Balboni, 2019).

Di seguito si presentano le diverse fasi in cui si articola la proposta.

Partendo dall'analisi dei bisogni linguistici degli apprendenti e del contesto di apprendimento, la prima fase prevede la definizione del livello linguistico di riferimento su cui si lavorerà, aspetto fondamentale per stabilire i criteri di progettazione del materiale didattico e per orientare lo sviluppo dell'UdA.

La seconda fase è dedicata alla selezione del testo di partenza, preferibilmente un testo autentico in lingua italiana (divulgativo, disciplinare o enciclopedico), disponibile in formato digitale o digitalizzato tramite strumenti Riconoscimento Ottico dei Caratteri (*Optical Character Recognition*, d'ora in poi OCR).

Segue l'analisi del livello linguistico (fase 3), finalizzata a collocare il testo scelto entro uno dei sei livelli di competenza del QCER. Per la determinazione del livello linguistico dell'italiano, un valido strumento di analisi della complessità testuale è MALT\_IT2<sup>9</sup> (Forti et al., 2020) o, in alternativa, servizi di intelligenza artificiale impiegati in modo critico e guidato. Per una più precisa determinazione delle strutture morfosintattiche e del lessico di riferimento del livello individuato è, inoltre, possibile avvalersi di strumenti come sillabi e referenziali linguistici. In particolare, per l'italiano risultano fondamentali le indicazioni fornite da *Il Profilo della lingua italiana* (Spinelli, Parizzi 2010), il quale offre un quadro di riferimento per la selezione del lessico, grazie alle liste lessicali organizzate per aree semantiche e livelli linguistici. Altrettanto utili risultano essere i principali sillabi per l'insegnamento dell'italiano come L2 (Lo

<sup>5</sup> OmegaT. Url: <https://omegat.org/> (cons. il 05/02/2026)

<sup>6</sup> Open source. Url: <https://opensource.org/osd> (cons. il 05/02/2026)

<sup>7</sup> Guida per l'utente OmegaT. Url: [https://hackmd.io/@omegat/for\\_cat\\_beginners](https://hackmd.io/@omegat/for_cat_beginners) (cons. il 05/02/2026)

<sup>8</sup> Unicode. Url: <https://home.unicode.org/> (cons. il 05/02/2026)

<sup>9</sup> MALT\_IT2. Url: è disponibile: <https://lol.unistrapg.it/malt/> (cons. il 06/02/2026)

Duca, 2006; Benucci, 2007), che presentano un catalogo, indicativo e non esaustivo, di forme per le diverse funzioni comunicative e delle strutture grammaticali pertinenti ai sei livelli di competenza. Si procede poi, nella quarta fase, con un lavoro preliminare di adattamento del testo di partenza al livello linguistico stabilito nella fase 1. Si interviene sulla struttura sintattica, morfosintattica e lessicale del testo, applicando strategie di semplificazione, riscrittura controllata e facilitazione, in funzione dei destinatari (età e livello linguistico) e degli obiettivi didattici prefissati stabiliti. In una quinta fase, il testo viene sottoposto a un processo di TA dall'italiano al sardo mediante Apertium italiano-sardo (Fronteddu et al., 2017; Tyers et al., 2017), integrato con l'impiego della TAO, OmegaT. Segue, nella fase 6, il post-editing (Vedi paragrafo 3), durante il quale non ci si limita alla correzione formale dell'output della traduzione automatica, ma si applica una riflessione di natura glottodidattica e acquisizionale analoga a quella adottata nella fase di adattamento del testo italiano. In particolare, la revisione tiene conto non solo della correttezza morfosintattica, ma anche dell'adeguatezza delle scelte lessicali e strutturali rispetto al livello linguistico precedentemente stabilito. In assenza, allo stato attuale, di sillabi e referenziali per la lingua sarda,<sup>10</sup> risulta utile far riferimento in modo comparativo alle risorse disponibili per le principali lingue romanze affini, come il *Plan curricular del Instituto Cervantes* (Instituto Cervantes 2007) per lo spagnolo e i *Référentiels pour le français* (Beacco et al. 2004, Beacco et al. 2007, Beacco et al. 2008, Beacco et al. 2011, Riba 2016). La fase 7, grazie all'uso di OmegaT, consente la costruzione progressiva di memorie di traduzione e glossari terminologici disciplinari, che possono essere riutilizzati in successive UdA o condivisi per altri progetti. Tale processo contribuisce non solo alla coerenza linguistica dei materiali prodotti, ma anche alla creazione di risorse digitali per il sardo, rafforzandone la presenza nello spazio educativo e rispondendo agli obiettivi di valorizzazione e normalizzazione linguistica perseguiti dalle politiche regionali. In questa prospettiva, l'integrazione tra CLIL, TA e TAO si configura come un impianto metodologico orientato all'efficienza produttiva, in cui il *post-editing guidato* consente una significativa riduzione dei tempi di elaborazione dei materiali didattici. Qualora la metodologia venga adottata come supporto alla produzione di materiali didattici per il sardo su larga scala, le memorie di traduzione in formato TMX generate attraverso l'uso di OmegaT potranno essere ri-utilizzate e condivise in progetti successivi, favorendo una maggiore coerenza terminologica e testuale. Tali risorse potranno inoltre costituire basi dati utili per l'addestramento di sistemi di IA e di motori di NMT, contribuendo allo sviluppo progressivo dell'ecosistema digitale della lingua sarda.

## 6. "SA PESADA DE SU FASCISMU": UN CASO DI STUDIO

Come caso di studio utile a mostrare un esempio applicativo di questa metodologia, si propone il modello pratico condiviso con i corsisti del Master di II livello in Didattica della Lingua Sarda. Si tratta di UdA di storia, svolta in modalità CLIL, rivolta a studenti del terzo anno della scuola secondaria di primo grado. L'UdA è intitolata *Sa Pesada de su Fascismu* ("L'ascesa del Fascismo"). Coerentemente con l'impianto metodologico sopra delineato, la fase iniziale del lavoro ha previsto la definizione del livello di riferimento e la selezione dei testi funzionali alla progettazione dell'UdA. A tal fine, sono stati individuati testi espositivi di tema storico tratti dal terzo volume del manuale *Fino a noi* (Fossati et al., 2020), attualmente in adozione nella scuola secondaria di primo grado. I testi sono stati successivamente digitalizzati mediante strumenti OCR per consentirne l'elaborazione didattica e traduttiva. Il livello linguistico stabilito prendendo in considerazione l'età e il grado scolastico frequentato dai potenziali destinatari dell'UdA (12-13 anni, scuola secondaria di primo grado) è il B1 del QCER. Lo strumento utilizzato nella terza fase per determinare il livello dei testi secondo il QCER è la piattaforma MALT\_IT2 (Forti et al., 2020). I testi sono risultati essere prevalentemente di livello B1, sebbene presentino anche tratti lessicali e strutturali riconducibili al livello B2 e, in misura marginale, al livello C1. Per questo motivo si è reso necessario un lavoro di semplificazione e facilitazione che si è limitato prevalentemente a interventi di linearizzazione e chiarificazione della struttura testuale e di

---

<sup>10</sup> Per quanto ad oggi il sardo non disponga di sillabi e referenziali, analogamente a quanto si osserva per altre lingue romanze, si segnala, tuttavia come nell'ambito dei progetti PeDESS *Pedagogia e Didattica per l'Educazione linguistica nella Scuola e nella comunità Sarda* (finanziato dall'Unione europea su fondi NextGenerationEU) e PLUREs: *PLUringuismo in Rete* (finanziato nell'ambito dell'alleanza "European Digital UniverCity" - EDUC) del Dipartimento di Lettere, Lingue e Beni Culturali dell'Università degli Studi di Cagliari sia stata avviata un'attività di ricerca finalizzata proprio alla loro progressiva elaborazione di questi strumenti (Deiana et al., 2025).

semplificazione di alcune scelte lessicali che non risultavano pertinenti per il livello stabilito per l'UdA ossia il B1.

Si è poi proceduto alla creazione di un nuovo progetto di traduzione su OmegaT, nel quale il testo è stato segmentato in unità di traduzione e successivamente sottoposto a un processo di TA su Apertium ita-srd, disponibile nella piattaforma tramite *plug-in*. Come illustrato nella Tabella 1,<sup>11</sup> l'*output* di Apertium è di buona qualità e il processo di *post-editing* sul prodotto della TA è legato soprattutto a scelte stilistiche o a eventuali vuoti traduttologici e ambiguità, tipicamente associati a un traduttore *rule-based* di tipo superficiale qual è Apertium (vedi paragrafo 3).

<b>Source Text</b>	<b>Apertium ita-srd</b>	<b>Post-editing</b>
La nascita del fascismo in Italia ha avuto luogo nel clima di crisi politica, sociale ed economica che seguì la Prima guerra mondiale. Alla fine del conflitto, il Paese si trovò attraversato da profonde tensioni.	<u>Sa nàschida de su fascismu in Itàlia at tentu logu</u> in su clima de crisi polìtica, sotziale e econòmica <u>chi aiat sighidu sa</u> Prima gherra mundiale. <u>A sa fine</u> de sa cuntierra, <u>su Paisu si fiat agatadu rugradu</u> dae profundas tensiones.	<b>Su fascismu est nàschidu in Itàlia</b> in su clima de crisi polìtica, sotziale e econòmica <b>a pustis de</b> sa Prima gherra mundiale. <b>A s'agabbu</b> de sa cuntierra, <b>in su Paisu b'aiat àpidu abolotu mannu.</b>

**Tabella 1. Un estratto del testo fonte e le versioni corrispondenti di Apertium e *post-editing*.**

Dal progetto di traduzione è stato ricavata una memoria di traduzione in formato TMX di 184 segmenti con un glossario in formato CSV di 10 espressioni legate alla storia del fascismo (es. it. "fasci di combattimento" → srd. "fàscios de cumbata"). La memoria è stata poi esportata e distribuita ai consisti del Master per la traduzione di testi finalizzati alla creazione di altre tre UdA, sviluppate su tematiche storiche e rivolte allo stesso target di discenti, con un risparmio di tempo stimato del 40%.

La memoria di traduzione, integrata dagli aggiornamenti derivati dai successivi progetti di traduzione, conta un totale di 344 segmenti, e un incremento del glossario a 23 espressioni.

Per quanto concerne la metodologia di valutazione del prodotto della TA in riferimento al *post-editing*, sono state utilizzate metriche automatiche basate sulla distanza testuale già utilizzate per la valutazione preliminare di Apertium ita-srd (Tyers et. al, 2017; Fronteddu et. al 2017). Nello specifico si è fatto ricorso al *Word Error Rate* (WER), che misura il numero minimo di operazioni (sostituzioni, inserzioni e cancellazioni) necessarie per trasformare l'*output* della traduzione automatica nel testo di riferimento, normalizzate sulla lunghezza di quest'ultimo (Lee et al., 2023) e al *Position-independent Error Rate* (PER) (Nießen et al., 2000), che valuta la sovrapposizione lessicale tra i due testi senza considerare l'ordine delle parole. Si è rivelato particolarmente utile in presenza di variazioni sintattiche; infatti una valutazione preliminare su un campione di 344 segmenti ha evidenziato un WER relativamente elevato (≈50%), dovuto principalmente a interventi di riformulazione stilistica (vedi sopra), mentre il PER risulta significativamente inferiore (≈30%), indicando una buona copertura lessicale da parte del sistema. Ciò conferma come l'*output* di Apertium costituisca una base adeguata per il *post-editing* in ambito didattico.

## 7. CONCLUSIONE

L'esperienza descritta mostra come l'integrazione tra CLIL, TA e TAO e competenze glottodidattiche possa costituire una strategia efficace per la produzione di risorse didattiche in sardo. Tale organizzazione consente di ottimizzare significativamente i tempi di produzione, grazie alla generazione automatica di bozze e al riuso sistematico delle risorse, concentrando l'intervento umano nella fase di revisione, fase che può avvalersi anche degli strumenti e delle risorse messe a disposizione dalla glottodidattica moderna. In questa prospettiva, le memorie di traduzione (TMX) e i glossari costruiti durante il processo risultano riutilizzabili e condivisibili in contesti successivi, favorendo coerenza terminologica e accelerando la produzione di nuovi materiali. Inoltre, tali risorse possono costituire basi dati utili per l'addestramento di sistemi di IA e NMT, contribuendo allo sviluppo progressivo dell'ecosistema digitale della lingua sarda.

<sup>11</sup> Nella seconda colonna relativa all'*output* di Apertium, alcune porzioni di testo sono evidenziate tramite sottolineatura per segnalare i segmenti oggetto di intervento. Tali elementi corrispondono alle forme riportate in grassetto nella colonna del *post-editing*, dove sono state modificate per migliorarne adeguatezza linguistica e scorrevolezza.

La metodologia qui presentata consente quindi non solo di adattare materiali esistenti, ma anche di produrre in tempi rapidi nuove risorse, contribuendo allo sviluppo dell'offerta didattica che per il sardo è ancora dotata di pochi strumenti.

Come sviluppo futuro, è auspicabile l'estensione del modello attraverso studi di caso fondati su sperimentazioni didattiche in contesti reali di classe, che consentano non solo di verificarne l'efficacia operativa, ma anche di raccogliere evidenze empiriche e *feedback* da parte dei docenti in merito all'utilità e alla funzionalità percepita della metodologia proposta.

## RINGRAZIAMENTI

Ringraziamo Unica per questa opportunità. Per ragioni di trasparenza, comunichiamo di aver fatto ricorso all'uso di supporti IA (ChatGPT), per analisi dei dati e formattazione delle citazioni.

Ringraziamo il gruppo di ricerca che ha coordinato il Master di II livello in Didattica della Lingua Sarda, con particolare riconoscimento alla prof.ssa Giulia Murgia.

## BIBLIOGRAFIA

- Alsaawi, A. M. (2025). Impact of Computer-Aided Translation Tools on the Product of Future Translators. *Journal of Language Teaching and Research*, 16(5), 1738–1748. <https://doi.org/10.17507/jltr.1605.31>
- Balboni, P. E. (2019). *Le sfide di Babele. Insegnare le lingue nelle società complesse*. Quarta edizione. Torino: UTET.
- Balling, L. W., Carl, M., & O'Brian, S. (2014). *Post-editing of machine translation: Processes and applications*. Cambridge Scholars Publishing. <https://books.google.com/books?hl=en&lr=&id=W3IxBwAAQBAJ&oi=fnd&pg=PR5&dq=info:5WiD9E8QKKoJ:scholar.google.com&ots=fallTtwuBvj&sig=d22N0WiK49EkhdaHf2Cvqeg7ezo>
- Beacco, J. C., Bouquet, S., Porquier R. (2004). *Un référentiel. Niveau B2 pour le français*. Parigi. Didier.
- Beacco, J. C., Porquier R. (2007). *Un référentiel. Niveau A1 pour le français*. Parigi. Didier.
- Beacco, J. C., Lepage, S., Porquier R., Riba, P. (2008). *Un référentiel. Niveau A2 pour le français*. Parigi. Didier.
- Beacco, J. C., Blin, B., Houles, E., Lepage, S., Riba, P. (2011). *Un référentiel. Niveau B1 pour le français*. Parigi. Didier.
- Benucci, A. (2007). *Sillabo di italiano per stranieri. Una proposta del Centro Linguistico dell'Università per Stranieri di Siena*. Perugia, Guerra Edizioni.
- Castilho, S., Moorkens, J., Gaspari, F., Calixto, I., Tinsley, J., & Way, A. (2017). Is Neural Machine Translation the New State of the Art? *The Prague Bulletin of Mathematical Linguistics*, 108(1), 109–120. <https://doi.org/10.1515/pralin-2017-0013>
- Consiglio d'Europa (2002). *Quadro comune europeo di riferimento per le lingue: apprendimento, insegnamento, valutazione*. Firenze. La Nuova Italia-Oxford.
- Consiglio d'Europa (2020). *Quadro comune europeo di riferimento per le lingue: apprendimento, insegnamento, valutazione. volume complementare. Italiano LinguaDue*, 12(2). DOI: <https://doi.org/10.13130/2037-3597/15120>
- Coyle, D., Hood, P., & Marsh, D. (2010). *Content and Language Integrated Learning*. Cambridge University chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/[https://assets.cambridge.org/97805211/30219/excerpt/9780521130219\\_excerpt.pdf](https://assets.cambridge.org/97805211/30219/excerpt/9780521130219_excerpt.pdf)
- Deiana, I., Grosso, G. I., Marra, A., Murgia, G., Puddu, N. (2025). *L'insegnamento della lingua sarda Strumenti e proposte didattiche. Presentazione al workshop L'insegnamento delle lingue minoritarie in contesti educativi: pratiche, problemi, prospettive. LVIII Congresso internazionale SLI. Grammaticalizzazione. 11-13 settembre 2025. L'Aquila*.
- Elhamayed, S. A., & Nour, M. (2025). Overview of deep learning and large language models in machine translation: A special perspective on the Arabic language. *Journal of Electrical Systems and Information Technology*, 12(1), 27. <https://doi.org/10.1186/s43067-025-00211-2>
- Forti, L., Grego Bolli, G., Santarelli, F., Santucci, V., & Spina, S. (2020). MALT-IT2: A New Resource to Measure Text Difficulty in Light of CEFR Levels for Italian L2 Learning. In N. Calzolari, F. Béchet, P. Blache, K. Choukri, C. Cieri, T. Declerck, S. Goggi, H. Isahara, B. Maegaard, J. Mariani, H. Mazo, A. Moreno, J. Odijk, & S. Piperidis (A c. Di), *Proceedings of the Twelfth Language Resources and*

- Evaluation Conference (pp. 7204–7211). European Language Resources Association.  
<https://aclanthology.org/2020.lrec-1.890/>
- Fossati, L., Rossi, M., & Bianchi, P. (2020). *Fino a noi. Storia, società e cittadinanza*. Milano: Zanichelli.
- Fronteddu, G., Font, H. A. i, & Tyers, F. M. (2017). Machine translation from Catalan to Sardinian: A translation tool for a language in the process of standardisation. *Linguamática*, 9(2), Articolo 2.  
<https://doi.org/10.21814/lm.9.2.255>
- Grosso, G. I., Marra, A., Murgia, G. (2024). Certifying a minority language: a first description of the Sardinian case. *MOSAIC*, 15, 105-119. DOI: 10.69117/MOSAIC.15.1.24.006
- Haddow, B., Bawden, R., Miceli Barone, A. V., Helcl, J., & Birch, A. (2022). Survey of Low-Resource Machine Translation. *Computational Linguistics*, 48(3), 673–732.  
[https://doi.org/10.1162/coli\\_a\\_00446](https://doi.org/10.1162/coli_a_00446)
- Instituto Cervantes (2007). Plan curricular del Instituto Cervantes. Niveles de referencia para el español. Madrid, Edelsa | Editorial Español lengua extranjera.  
[https://cvc.cervantes.es/ensenanza/biblioteca\\_ele/plan\\_curricular/](https://cvc.cervantes.es/ensenanza/biblioteca_ele/plan_curricular/)
- Kapusta, J., Benko, L., Munkova, D., & Munk, M. (2021). Analysis of Edit Operations for Post-editing Systems. *International Journal of Computational Intelligence Systems*, 14(1), 1.  
<https://doi.org/10.1007/s44196-021-00048-3>
- Kenny, D. (2022). Machine translation for everyone: Empowering users in the age of artificial intelligence. <https://doi.org/10.5281/ZENODO.6653406>
- Koponen, M., Salmi, L., & Nikulin, M. (2019). A product and process analysis of post-editor corrections on neural, statistical and rule-based machine translation output. *Machine Translation*, 33(1), 61–90.  
<https://doi.org/10.1007/s10590-019-09228-7>
- Lakew, S. M., Erofeeva, A., Negri, M., Federico, M., & Turchi, M. (2018). Transfer Learning in Multilingual Neural Machine Translation with Dynamic Vocabulary. In M. Turchi, J. Niehues, & M. Federico (A c. Di), *Proceedings of the 15th International Conference on Spoken Language Translation* (pp. 54–61). International Conference on Spoken Language Translation.  
<https://doi.org/10.48550/arXiv.1811.01064>
- Lee, S. et al. (2023). A Survey on Evaluation Metrics for Machine Translation. *Mathematics*, 11(4), 1006.
- Lo Duca, M. G. (2006). *Sillabo di italiano L2*. Roma, Carocci.
- Marra, A. (2021). L'insegnamento della lingua sarda. In Luise, M. C., Vicario, F. (A c. Di.), *Le lingue regionali a scuola. Competenze e certificazione didattica*. UTET, Torino, 197-224.
- Martín-Mor, A., Sánchez-Gijón, P., & Huerta, R. P. i. (2016). *Tradumàtica: Tecnologies de la traducció*. Eumo Editorial.
- Mattioda, M. M., Molino, A., & Cinato, L. (2023). L'intelligenza artificiale per la traduzione: Orizzonti, pratiche e percorsi formativi. *mediAzioni*, 39, A1–A16. <https://doi.org/10.6092/issn.1974-4382/18784>
- Micali, I. (2023). Quale didattica per le lingue minoritarie? Approcci e modelli per un'educazione plurilingue. *Italiano LinguaDue*, 15/1, 730-746.
- Mereu, D. (2024). Una prospettiva ecologica allo studio delle lingue di minoranza: Il sardo e le sue varietà. *PRT*. <https://iris.unito.it/handle/2318/1988912>
- Monti, J. (2023). Per una didattica della traduzione automatica. *mediAzioni*, 39, A55–A83.  
<https://doi.org/10.6092/issn.1974-4382/18787>
- Nießen, S., Och, F. J., Leusch, G., & Ney, H. (2000). An evaluation tool for machine translation: Fast evaluation for MT research. In *Proceedings of the Second International Conference on Language Resources and Evaluation (LREC 2000)*. Athens.
- Pombal, J., Guerreiro, N. M., Rei, R., & Martins, A. F. T. (2025). Adding Chocolate to Mint: Mitigating Metric Interference in Machine Translation (arXiv:2503.08327). *arXiv*.  
<https://doi.org/10.48550/arXiv.2503.08327>
- Regione Autonoma della Sardegna. (2018). Legge regionale 3 luglio 2018, n. 22 — Disciplina della politica linguistica regionale (B.U. Regione Sardegna 05/07/2018, n. 32). Testo vigente sulla tutela, promozione e diffusione delle lingue storiche, inclusa la lingua sarda.  
<https://leggi regionali.regione.sardegna.it/legge-regionale?data=3-7-2018&numero=22>

- Regione Autonoma della Sardegna. (2006). *Limba Sarda Comuna — Norme linguistiche di riferimento per la lingua sarda*. Deliberazione adottata dalla Giunta regionale (norma ortografica ufficiale per la lingua sarda). [https://www.regione.sardegna.it/documenti/1\\_72\\_20060418160308.pdf](https://www.regione.sardegna.it/documenti/1_72_20060418160308.pdf)
- Regione Autonoma della Sardegna. (2025), *PIANO DI POLITICA LINGUISTICA 2025-2029 - RELAZIONE*. <https://delibere.regione.sardegna.it/delibera/116675>
- Riba, P. (2016). *Niveau C1/C2 pour le français. Éléments pour un référentiel*. Parigi. Didier.
- Riediger, H., & Galati, G. (2023). *La traduzione nell'era dell'IA: Nuovi ruoli, nuove competenze, nuova formazione*. *mediAzioni*, 39, A35–A54. <https://doi.org/10.6092/issn.1974-4382/18786>
- Shormani, M. Q. (2024). *Generative linguistics contribution to artificial intelligence: Where this contribution lies?* (arXiv:2410.20221). arXiv. <https://doi.org/10.48550/arXiv.2410.20221>
- Spinelli, B., Parizzi, F. (2010). *Profilo della lingua italiana. Livelli di riferimento del QCER A1, A2, B1, B2*. Firenze, La Nuova Italia.
- Stato Italiano. (1999). *Legge 15 dicembre 1999, n. 482 – Norme in materia di tutela delle minoranze linguistiche storiche (G.U. n. 297, 20 dicembre 1999)*. <https://www.mim.gov.it/lingue-di-minoranza>
- Terribile, S. (2024). *Is post-editing really faster than human translation?* *Translation Spaces*, 13(2), 171–199. <https://doi.org/10.1075/ts.22044.ter>
- Tyers, F. M., Alòs i Font, H., Fronteddu, G., & Martín-Mor, A. (2017). *Rule-Based Machine Translation for the Italian–Sardinian Language Pair*. *The Prague Bulletin of Mathematical Linguistics*, 108(1), 221–232. <https://doi.org/10.1515/pralin-2017-0022>
- Wang Et Al. 2022 | PDF | Machine Learning | Receiver Operating Characteristic. (s.d.). Recuperato 6 febbraio 2026, da <https://it.scribd.com/document/882065912/Wang-et-al-2022>
- Wu, Y., Schuster, M., Chen, Z., Le, Q. V., Norouzi, M., Macherey, W., Krikun, M., Cao, Y., Gao, Q., Macherey, K., Klingner, J., Shah, A., Johnson, M., Liu, X., Kaiser, Ł., Gouws, S., Kato, Y., Kudo, T., Kazawa, H., ... Dean, J. (2016). *Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation* (arXiv:1609.08144). arXiv. <https://doi.org/10.48550/arXiv.1609.08144>