



ALLEGATO A

SELEZIONE PUBBLICA PER TITOLI E COLLOQUIO PER IL CONFERIMENTO DI N. 1 ASSEGNO EXPERIENCED DI 24 MESI PER LO SVOLGIMENTO DI ATTIVITA' DI RICERCA AI SENSI DELL'ART. 22 DELLA LEGGE N. 240/2010 PRESSO IL DIPARTIMENTO DI LINGUE, LETTERATURE E CULTURE STRANIERE SSD L-LIN/01 - GLOTTOLOGIA E LINGUISTICA (SC 10/G1 - GLOTTOLOGIA E LINGUISTICA) TIPO B - CODICE PICA: 21AR009

PROGETTO DI RICERCA

"Analisi sociolinguistica e trattamento informatico dei papiri Tjäder"

Struttura di ricerca: Dipartimento di Lingue, letterature e culture straniere

Durata dell'assegno: 24 mesi

Area scientifica: 10 - Scienze dell'antichità, filologico-letterarie e storico-artistiche

Settore concorsuale: 10/G1 - Glottologia e Linguistica

Settore scientifico disciplinare: L-LIN/01 - Glottologia e Linguistica

Responsabile scientifico: Prof. Pierluigi Cuzzolin

1) Obiettivo:

Il progetto intende offrire una dettagliata analisi linguistica dei papiri documentari italiani redatti tra il 445 e il 700 in area ravennate, e editi dal filologo e latinista svedese Jan-Olof Tjäder. Dal punto di vista metodologico, lo studio si propone di integrare gli approcci e i metodi della linguistica e sociolinguistica storica con quelli sviluppati nell'ambito delle *digital humanities* applicate alle lingue antiche, con particolare riguardo al trattamento automatico e all'annotazione linguistica dei testi e alla cosiddetta estrazione delle informazioni.

Il corpus prescelto, oltre a costituire una fonte storica di primario rilievo, rappresenta una miniera di dati linguistici e sociolinguistici non ancora esplorata con sistematicità. In particolare, i papiri ravennati consentono studi di dettaglio sul tema del contatto linguistico, in special modo tra il latino, la lingua parlata dalla maggior parte della popolazione, e il greco, esito della costituzione dell'esarcato bizantino intorno all'area ravennate (VI-VIII secolo). L'obiettivo principale del progetto è contribuire ad una più approfondita conoscenza della variabilità linguistica nel mondo latinofono tra il V e il VII secolo, con particolare riguardo alle differenze sociali e culturali osservabili, in una moderna prospettiva di sociolinguistica storica.

Il progetto si articolerà in due fasi. La prima fase sarà dedicata all'analisi preliminare delle caratteristiche linguistiche dei testi, al fine di individuare i fenomeni di natura grafica e fonetica, morfologica, sintattica e lessicale più rilevanti. Per la parte informatica, si procederà ad una revisione dell'edizione digitale già disponibile dei papiri (nel sito: papyri.info), ad una implementazione dei metadati e alla processazione dei testi per l'annotazione linguistica automatica con strumenti *open-source*, a cui seguirà una revisione manuale. La seconda fase prevederà un'annotazione aggiuntiva dei testi per le specifiche finalità del progetto, con particolare riguardo alle informazioni di natura sociolinguistica. Si svilupperanno, infine, delle sintesi descrittive e interpretative delle caratteristiche linguistiche, testuali e sociolinguistiche di una significativa selezione di documenti, anche nell'ottica di elaborare confronti con fonti documentarie di altre aree della penisola italiana coeve o più tarde, per esempio i documenti già editi di Cava de' Tirreni. Di questa selezione di testi verrà offerta anche la prima traduzione in lingua italiana.

2) Definizione dell'ambito della ricerca:

I papiri ravennati rappresentano un'importante fonte di dati per lo studio delle trasformazioni linguistiche e culturali del latino tra tardo antico e alto medioevo. Fra queste ci sono volgarismi ed elementi anticipatori di caratteristiche strutturali che troveranno continuazione nei volgari romanzi. Come è noto, tra il V e il VII secolo l'area intorno a Ravenna si caratterizzò per un forte plurilinguismo. Popolazioni di varia provenienza si trovarono ad usare codici linguistici diversi, esibendo, come gli



stessi papiri documentano, competenze linguistiche differenziate. In questa prospettiva, di particolare interesse è l'analisi del latino messo per iscritto da parlanti per i quali il latino non era la lingua madre e il confronto sistematico con le produzioni dei parlanti nativi. Vanno tenute presenti anche le complesse questioni sollevate dal rapporto tra scritto e parlato e la permeabilità dello scritto a fenomeni della lingua parlata. Da un punto di vista teorico e metodologico, tali analisi consentono di testare l'efficacia descrittiva e interpretativa dei modelli della variazione e del cambiamento linguistico in contesti multilingui elaborati dalla sociolinguistica, soprattutto a partire dall'osservazione di situazioni moderne. La proiezione di questi modelli su situazioni linguistiche del passato non può essere però immediata: è centrale tenere conto delle difficoltà e dei rischi di associare in maniera diretta le differenze linguistiche e stilistiche dei testi alle caratteristiche sociolinguistiche degli scriventi. Gli indizi linguistici, spesso minuti, devono essere valutati con cautela alla luce del più generale contesto storico e socio-culturale in cui i documenti furono prodotti e di tutte le informazioni che i testi stessi forniscono. In questa prospettiva, le dettagliate informazioni di natura paleografica di cui l'edizione dei papiri ravennati curata da Tjäder è ricca, unite alla considerazione delle diverse tipologie documentarie, risultano di particolare importanza.

Oltre che per i modelli della sociolinguistica storica, il progetto si propone come banco di prova anche per le potenzialità degli strumenti informatici nell'analisi di testi, che presentano notevoli differenze di registro e di stile. La digitalizzazione già esistente dei papiri italiani di area ravennate ad opera di *The Duke Databank of Documentary Papyri*, disponibile nel sito papyri.info sotto la licenza "Creative Commons Attribution 3.0 License", rappresenta infatti una risorsa preziosa da valorizzare e implementare. Quanto mai attuale si rivela il riuso delle risorse digitali già disponibili, ma anche il favorire la confrontabilità tra i dati che provengono dagli ormai numerosi corpora digitali, l'accessibilità e la condivisione dei risultati in contesto accademico e scientifico, la loro ricercabilità e il loro ulteriore riuso. Anche in questo caso, l'uso degli strumenti per l'annotazione linguistica dei testi sviluppati per le lingue antiche non potrà essere immediato. Il latino dei papiri italiani di area ravennate presenta infatti numerose varianti di natura grafica e morfologica rispetto alla norma classica sulla quale gli strumenti sono basati. Ciò rischia di compromettere la correttezza delle annotazioni automatiche. Soltanto un successivo intervento manuale potrà consentire di valutare la percentuale di precisione delle annotazioni e la quantità e tipologia di interventi necessari per migliorare l'analisi automatica di testi latini che si discostano dalla norma classica spesso in maniera considerevole.

3) Descrizione del progetto:

Il progetto si articola in due fasi. La prima fase sarà dedicata all'analisi preliminare delle caratteristiche linguistiche dei testi, al fine di individuare i fenomeni di natura grafica e fonetica, morfologica, sintattica e lessicale più rilevanti. Per la parte informatica, si procederà ad una revisione dell'edizione digitale disponibile e ad una implementazione dei metadati, annotando la tipologia documentaria di ciascun testo e le relative formule caratterizzanti. La codifica sarà effettuata secondo lo standard XML TEI P5.

Mediante l'uso di strumenti open-source, si procederà poi ad una processazione automatica dei testi che fornirà una prima lemmatizzazione e analisi morfologica e sintattica dei documenti. Considerata la particolare natura dei testi, la loro frammentarietà e le varianti, in particolare di natura grafica e morfologica, rispetto alla norma classica, sulla quale gli strumenti informatici sono addestrati, sarà necessaria una revisione e correzione manuale delle annotazioni automatiche.

La seconda fase prevede un'annotazione aggiuntiva dei testi secondo gli scopi del progetto, con particolare riguardo ai dati sociolinguistici. Saranno annotati antroponomi e toponimi e si effettuerà una lemmatizzazione di secondo livello per aree semantiche (termini di parentela, termini giuridici, cultura materiale, ecc.). Saranno infine annotate le realizzazioni delle strutture argomentali dei verbi. L'operazione di annotazione consentirà di ottenere un corpus digitale ricercabile dei papiri ravennati da cui sarà possibile estrarre informazioni linguistiche di natura grafica, fonetica, morfologica, sintattica e lessicale, nonché di natura sociolinguistica. L'annotazione linguistica dei testi, insieme ad un'analisi qualitativa di dettaglio, consentirà l'elaborazione di sintesi descrittive e interpretative delle caratteristiche linguistiche, sociolinguistiche e testuali di una significativa selezione di documenti, che saranno poi tradotti. Tali sintesi costituiranno la base per sviluppare confronti con fonti documentarie di altre aree della penisola italiana coeve o più tarde, per esempio i documenti già editi di Cava de' Tirreni.



Parole chiave: Sociolinguistica storica, trattamento informatico dei dati, latino, lingue in contatto, fonti documentarie, traduzione dei linguaggi tecnici

4) Tempi di realizzazione e calendario delle attività:

Il progetto si articola in due anni, così ripartiti:

- 1) Revisione dell'edizione digitale dei papiri; implementazione dei metadati (tipologie e formule documentarie); analisi preliminare delle caratteristiche linguistiche dei singoli documenti e individuazione dei fenomeni linguistici (grafici e fonetici, morfologici, sintattici e lessicali) più significativi; processazione automatica dei testi (lemmatizzazione, analisi morfologica e sintattica) con strumenti open-source, ad esempio i progetti ospitati da *Perseus Digital Library* (Tufts University) o LEMLAT, e successiva revisione manuale.
- 2) Annotazione linguistica specifica per il corpus in XML TEI P5 (antroponimi, toponimi, lemmatizzazione di secondo livello per aree semantiche, struttura argomentale dei verbi); elaborazione di sintesi descrittive e interpretative delle caratteristiche linguistiche, sociolinguistiche e testuali di una significativa selezione di documenti; traduzione della selezione di testi in italiano; confronto con i documenti del IX secolo di Cava de' Tirreni.

5) Originalità e ricaduta sull'avanzamento della conoscenza:

I papiri documentari italiani dell'area ravennate editi da Tjäder, pur riconosciuti come una fonte storica di grande importanza, non sono stati ancora oggetto di indagini linguistiche sistematiche. I dati che essi offrono possono contribuire ad una migliore comprensione:

- i) delle dinamiche di cambiamento linguistico del latino tra il V e VII secolo in contesto plurilingue;
- ii) di come persone che non avevano il latino come lingua madre percepissero il latino parlato;
- iii) della situazione sociolinguistica tardoantica in un'area come quella dell'esarcato bizantino di Ravenna che rappresenta un contesto paradigmatico di multilinguismo nel passato.

Da un punto di vista teorico e metodologico il progetto mira a riconsiderare in maniera critica:

- i) le possibili inferenze che analisi di testi documentari consentono sul rapporto tra abilità di lingua scritta e lingua parlata;
- ii) l'efficacia descrittiva e interpretativa dei modelli di variazione e cambiamento linguistico in situazioni di contatto elaborati dalla moderna sociolinguistica;
- iii) le potenzialità e i limiti degli strumenti informatici per il trattamento automatico delle lingue antiche quando applicati a testi che si allontanano, in maniera talvolta considerevole, dalle norme classiche.

Infine, va sottolineato che il progetto offre un contributo alla conoscenza complessiva della situazione linguistica della penisola italiana tra tardo antico e alto medioevo, concentrandosi su un'area e una cronologia per le quali sono necessarie ricerche di dettaglio, ponendo così le basi per utili comparazioni con altre fonti documentarie.

Iniziale bibliografia di riferimento

- Amelotti, M. (1975), «Il documento tabellionico tra l'Oriente e l'Italia», in Amelotti, M. / Costamagna, G., *Alle origini del notariato italiano*, Roma, Consiglio nazionale del notariato, pp. 49-68; 85-95.
- Bowern, C. / Evans, B. (eds.) (2015), *The Routledge Handbook of Historical Linguistics*, New York, Routledge.
- Cuzzolin, P. / Sornicola, R. (2018), «Papiro 37 Tjäder: una revisione sociolinguistica e testuale», in Bombi, R. / Costantini, F. (a cura di), *Percorsi linguistici e interlinguistici. Studi in onore di Vincenzo Orioles*, Udine, Forum, pp. 291-315.
- Hernández-Campoy, J. M. / Conde-Silvestre, J. C. (eds.) (2012), *The Handbook of Historical Sociolinguistics*, Malden / Oxford, Blackwell.
- Humbley, J. / Budin, G. / Laurén, C. (eds.) (2018), *Languages for Special Purposes: An International Handbook*, Berlino / Boston, De Gruyter.
- Janda, R. D / Joseph, B. D. / (eds.) (2003), *The Handbook of Historical Linguistics*, Oxford, Blackwell.
- Janda, R. D / Joseph, B. D. / Vance, B. S. (eds.) (2021), *The Handbook of Historical Linguistics*, Vol. II, Oxford, Blackwell.



**UNIVERSITÀ
DEGLI STUDI
DI BERGAMO**

Direzione personale
logistica
e approvvigionamenti

Gestione risorse umane
Gestione giuridica
e selezioni assegni di ricerca

Mauskopf Deliyannis, D. (2010), *Ravenna in late Antiquity*, Cambridge, Cambridge University Press.
Migliardi Zingale, L. (2008), «Sui papiri "ravennati", punto d'incontro tra Occidente ed Oriente: alcune riflessioni», *Aegyptus*, 88, pp. 149-163.
Tjäder, J.-O. (1955-1982), *Die nicht-literarischen lateinischen Papyri Italiens aus der Zeit 445-700*, voll. I-II, Svenska Institutet i Rom, Uppsala, Almqvist & Wiksell.
Ware, E. (2014), *The Social and Cultural Meaning of Names in Late Antique Italy, 313-604*, Master Theses, <http://scholarworks.wmich.edu/master_theses/510>.
Zanini, E. (1998), *Le Italie bizantine. Territorio, insediamenti ed economia nella provincia bizantina d'Italia (VI-VIII secolo)*, Bari, Edipuglia.