

Pietro G. Beltrami

(CNR - Opera del Vocabolario Italiano, Firenze, Italia)

*Il Tesoro della Lingua Italiana delle Origini*: caratteristiche, problemi, futuro.

La comunicazione intende presentare lo stato attuale del *Tesoro della Lingua Italiana delle Origini (TLIO)*, l'ormai ben noto vocabolario storico dell'italiano antico pubblicato in rete in corso d'opera ([www.vocabolario.org](http://www.vocabolario.org)), mettendone in evidenza le caratteristiche principali, alcuni aspetti problematici e le prospettive future, sia per quanto riguarda il vocabolario propriamente detto, sia per quanto riguarda le banche dati testuali su cui esso si fonda, consultabili allo stesso indirizzo.

Il *TLIO* è un progetto nato molto tempo fa, come parte di un vocabolario storico italiano che l'Accademia della Crusca iniziò nel 1965 e abbandonò nel 1983 al Consiglio Nazionale delle Ricerche, che già lo finanziava fin dal principio; da allora il lavoro è eseguito dall'Opera del Vocabolario Italiano, un Istituto del CNR con sede presso l'Accademia.

Il *TLIO* ha come fonte principale un corpus di testi anteriori alla fine del Trecento, scritti in tutte le varietà linguistiche italiane e disponibili in edizioni. In teoria il vocabolario dovrebbe utilizzare solo edizioni affidabili o riviste in funzione della redazione, ma in realtà per la completezza del lessico è necessario ricorrere anche a edizioni insoddisfacenti, così come si citano quando è necessario anche fonti esterne al corpus (questi problemi sono oggetto di una comunicazione separata nella sezione 17).

La prima versione del corpus utilizzabile per la redazione è stata completata alla fine del 1995, e conteneva circa 15 milioni di occorrenze; nel 1998 il corpus è stato reso interrogabile in rete in forma non lemmatizzata, in collaborazione con le Università americane di Chicago e di Notre Dame (accordo ItalNet); dal 2005 il corpus è interrogabile in rete in forma lemmatizzata. Dal 1998 per le operazioni in scrittura, inclusa la lemmatizzazione, e dal 2005 per l'accesso in rete, si usano procedure informatiche prodotte dall'Istituto (autore Domenico Iorio-Fili), cioè *GATTO* e *GattoWeb*. Precisamente, gli utenti hanno accesso a tre corpora: il *Corpus TLIO*, di 2000 testi (circa 21,85 milioni di occorrenze), lemmatizzato, che è quello che serve di base alla redazione del *TLIO*; il *Corpus TLIO aggiuntivo*, di 319 testi (circa 1,3 milioni di occorrenze) in attesa di lemmatizzazione; e il *Corpus OVI dell'Italiano antico*, in cui gli altri due corpora sono riuniti e resi consultabili insieme.

Il corpus è sempre in sviluppo, sia per quanto riguarda l'aggiunta di nuovi testi, sia per quanto riguarda la lemmatizzazione, e viene aggiornato con cadenza regolare ogni quattro mesi. Si può dire però che già contiene tutto l'essenziale per lo studio

dell'italiano antico, e, pur essendo stato costituito come uno strumento per la redazione del vocabolario, è diventato rapidamente, da quando è in rete, una risorsa di grande importanza per gli studiosi di linguistica italiana e romanza. Nella comunicazione congressuale si dirà di qualche problema relativo allo sviluppo, e in particolare di qualche aspetto della lemmatizzazione.

Le prime voci del *TLIO* sono state redatte nel 1996, e la pubblicazione in rete è cominciata nel 1997, con successivi cambiamenti e miglioramenti dell'interfaccia. Nel 2012 è stato superato il numero di 25.000 voci, che significa che il vocabolario è abbastanza vicino alla metà. Le norme di redazione si sono successivamente evolute e raffinate rispetto al primo disegno preparato nel 1996, con l'intento di contemperare l'accuratezza e la completezza di informazione e di analisi lessicale con le esigenze di avanzamento di un'opera di grandi dimensioni affrontata con mezzi relativamente modesti. La comunicazione intende soffermarsi su alcuni problemi teorici e pratici relativi alla redazione delle voci, di cui è possibile farsi un'idea preliminare sia dalla consultazione del vocabolario in rete, sia dalla lettura delle norme di redazione scaricabili integralmente dallo stesso sito.

*TLIO* = [www.vocabolario.org](http://www.vocabolario.org)

Pietro G. Beltrami, *La nuova lessicografia dell'italiano antico. Il Tesoro della Lingua Italiana delle Origini*, «Bollettino dell'Atlante Lessicale degli Antichi Volgari Italiani», I, 2008, pp. 33-52.

Pietro G. Beltrami, *The Lexicography of Early Italian: its Evolution and Recent Advances*, in *Perspectives on Lexicography in Italy and Europe*, ed. by Silvia Bruti, Roberta Cella and Marina Foschi Albert, Newcastle upon Tyne, Cambridge Scholars Publishing, 2009, pp. 27-53.