

Società Italiana di Filologia Romanza

*Presentazione di progetti di ricerca in corso
(Verona, 13 settembre 2014)*

1. Titolo progetto: *Tesoro della Lingua Italiana delle Origini (TLIO)*.
2. Durata prevista: al ritmo attuale di redazione (circa 2000 voci l'anno), sono previsti circa 15 anni per completare l'impresa.
3. Coordinatore: Lino Leonardi, direttore dell'Istituto CNR Opera del Vocabolario Italiano (OVI).
4. Partecipanti al gruppo di ricerca e articolazione di eventuali sottogruppi: 1) redazione TLIO all'OVI: Elena Artale, Maria Fortunato (assegnista UniSS, progetto ATLiSor), Mariafrancesca Giuliani, Pär Larson, Luca Morlino, Rossella Mosti, Sara Ravani (assegnista UniSS, progetto ATLiSor), Paolo Squillacioti, Zeno Verlati, Ilaria Zamuner (associata all'OVI); 2) gruppo di redazione DiVo (progetto OVI-Scuola Normale): Cosimo Burgassi, Diego Dotto, Elisa Guadagnini, Cristiano Lorenzi, Cristiano Lorenzi Biondi, Valentina Nieri, Giulio Vaccaro; 3) collaboratori tecnici OVI: Andrea Boccellari (informatica), Carmen Pasetto (amministrazione), Mario Malatesta (acquisizione testi); 4) gruppo di redazione di Notre Dame (USA): Pietro Bocchia, Lorenzo Dell'Oso, Leonardo Francalanci, Filippo Gianferrari, Annalia Marchisio, Paolo Pellicchia, Mary Prokop, Demetrio Yocum.
5. Finalità del progetto: Realizzazione del primo vocabolario storico dell'italiano, e in prima istanza della *tranche* medievale, dalle Origini (il testo più antico è l'*Indovinello veronese*) alla fine del XIV secolo (ma il testo più recente è l'edizione Negrone del volgarizzamento della Bibbia, fondata per gran parte su un incunabolo della seconda metà del XV secolo).

6. Metodologie applicate: Il vocabolario viene redatto di prima mano su un'ampia banca-dati, il *Corpus OVI dell'italiano antico* (oltre 23 milioni di occorrenze), peraltro aperto alla libera consultazione online, dal sito web dell'OVI (www.vocabolario.org) grazie al software lessicografico GattoWeb, anch'esso prodotto dell'Istituto. La redazione procede senza un'esame preliminare dei grandi dizionari di attestazioni dell'italiano (le cinque impressioni del *Vocabolario degli Accademici della Crusca*, il *Dizionario della Lingua Italiana* di Nicolò Tommaseo e Bernardo Bellini, il *Grande Dizionario della Lingua Italiana* fondato da Salvatore Battaglia), ma attingendo alla lemmatizzazione del *Corpus OVI*, che copre meno del 20% delle occorrenze ma che è stata realizzata con criteri funzionali alla redazione del *TLIO*. Il confronto con gli strumenti lessicografici avviene in un secondo momento, e ciò garantisce un'elaborazione personale del materiale da parte del redattore, che di per sé produce risultati originali. Il *Corpus OVI* annovera testi redatti in tutte le varietà italo-romanze entro fine del XIV secolo, molti dei quali mai spogliati in funzione lessicografica, cosicché il *TLIO* fornisce esempi, significati e voci mai registrati in un dizionario.

7. Risultati attesi: il *TLIO* conterà tra 60 e 70 mila voci, ma l'OVI mette a disposizione degli utenti l'intera documentazione del vocabolario (il *Corpus TLIO*) insieme con altre banche dati sviluppate a partire da materiali del corpus principale: *DiVo - Dizionario dei volgarizzamenti*, corpus dei volgarizzamenti medievali italiani di opere classiche e tardoantiche, *CLaVo - Classici Latini Volgarizzati*, corpus delle opere latine tradotte dai volgarizzamenti compresi nel Corpus DiVo, *ReMediA - Repertorio di Medicina Antica*, corpus di testi medico-scientifici nelle diverse lingue romanze, *ARTESIA - Archivio Testuale del Siciliano Antico*, corpus realizzato a cura di Mario Pagano presso l'Università di Catania, *Archivio Datini*, corpus delle lettere edite dell'Archivio del mercante Francesco di Marco Datini, elaborato dall'OVI per conto dell'Archivio di Stato di Prato. Altro corpus GATTO significativo è *LirIO - Lirica Italiana delle Origini*, raccolta della poesia italiana dalle Origini alla fine del Quattrocento, elaborata

sotto la direzione di Lino Leonardi dall'OVI, dalla Fondazione Ezio Franceschini onlus di Firenze e dall'Università di Siena; in attesa della prossima diffusione online con GattoWeb è disponibile in DVD: *LirIO. Corpus della lirica italiana delle origini in CD-ROM. 2. Dagli inizi al 1400*, a cura di Lino Leonardi e di Alessio Decaria, Pär Larson, Giuseppe Marrani, Paolo Squillacioti, Firenze, Edizioni del Galluzzo per la Fondazione Ezio Franceschini, 2013.

8. Stato di avanzamento della ricerca: attualmente (12.11.2014) sono disponibili online 28.924 voci, per lo più (87%) concentrate nel segmento alfabetico *A-N*, ma stante la pubblicazione online che non vincola alla stretta sequenza alfabetica sono già state diffuse voci che iniziano per *O* (154), *P* (1324), *Q* (108), *R* (283), *S* (1199), *T* (374), *U* (61), *V* (146), *X* (5), *Z* (79).
9. Esemplicazioni essenziali dei risultati finora raggiunti: le quasi 30 mila voci del *TLIO* coprono tutti gli àmbiti e le varietà della lingua italiana delle Origini: accanto alle numerose voci documentate in Toscana, voci attestate solo in milanese (per es. ATTASTARE v., o in perugino (BONITO s.m.), o in veneziano (CHEBA s.f.), ecc.; voci articolate e polisemiche (DIRITTO agg./avv./prep./s.m.) e voci con un'unica attestazione (EDIZIONE s.f.); voci auliche (FIORDALISO s.m./s.f.) e voci popolari (GRATTACACIO s.f.); voci misteriose, forse inesistenti (HUTRIC s.i.) e voci ben attestate (INTERPRETARE v.); voci del lessico gastronomico (LASAGNA s.f.), medico-veternario (MADDALEONE s.m.), tessile (NASPO s.m.), agricolo (OCCARE v.), architettonico (PALCO s.m.), e così via.