

SOMMARI DEGLI ARTICOLI IN ITALIANO E IN INGLESE

ILARIA ZAMUNER, Il glossario dell'«Antidotarium Nicolai» volgarizzato (ms. New Haven, Yale University, Historical Medical Library, 52, ff. 86v-96ra)

Dopo una breve introduzione sulla tradizione latina e romanza dell'*Antidotarium Nicolai*, viene proposto un glossario selettivo del volgarizzamento italo-romanzo più antico contenuto nel ms. New Haven, Yale University, Historical Medical Library, 52 (ff. 86v-96ra), latore di numerose prime attestazioni.

After a brief introduction on the Latin and Romance tradition of the *Antidotarium Nicolai*, there follows a selective glossary of the oldest vernacular Italian-Romance text found in the New Haven manuscript at Yale University, Historical Medical Library, 52 (ff.86v-96ra), the sources of many first instances.

NICOLETTA DELLA PENNA, La semantica di «immaginazione» nel medioevo italo-romanzo

Il contributo propone un'analisi semantica della parola *immaginazione* nelle sue attestazioni italo-romanze di epoca medioevale, fondata sullo studio della documentazione raccolta per la redazione di una voce per il *Tesoro della lingua italiana delle origini* (TLIO) e reperibile sui *corpora* dell'Opera del vocabolario italiano (OVI). L'indagine, condotta in prospettiva lessicologica, presenta la sfera semantica del termine attraverso i suoi principali nuclei di significato: questi, strettamente interconnessi per mezzo di estensioni, metonimie e usi metaforici, orbitano attorno ad un concetto che affonda le sue radici teoriche nella riflessione aristotelica elaborata nel *De anima* e tramandata alla cultura medioevale per mediazione dei medici-filosofi arabi. La differenza principale rispetto al sistema lessicale della lingua attuale risiede nel fatto che, nell'italiano antico, *immaginazione* può assumere anche un preciso valore tecnico e qualificarsi, dunque, come parola appartenente al lessico specifico di ambito filosofico.

This essay offers a semantic analysis of the word *immaginazione* in Italian-Romance instances dating back to the Middle Age, based on an analysis of the

documentation collected for drafting an entry for the *Tesoro della lingua italiana delle origini* (TLIO), that can be found in the *Corpora* of the Opera del Vocabolario Italiano (OVI). The research, conducted according to a lexicological approach, shows the semantic range of the word through its main nuclei of meaning: these, closely linked by means of extensions, metonymies and metaphorical usage, gravitate around a concept that is firmly anchored in Aristotelian thought as developed in *De anima* and transmitted to medieval culture through the role of Arab philosophers and doctors. The main difference from the lexical system of current language lies in the fact that, in ancient Italian, *immaginazione* can also assume a specific technical value and thus be considered as a word belonging to the specific lexicon of philosophy.

EUGENIO SALVATORE - GIUSEPPE ZARRA, «Partimoci di Firenze a di 10 agosto 1384». Lavoro filologico e lessicografico sui resoconti del viaggio in Terrasanta di Giorgio Gucci e Lionardo Frescobaldi

Il saggio prende in esame la vicenda filologico-editoriale e la ricezione lessicografica nel corso dell'Ottocento di due resoconti di un viaggio in Terrasanta compiuto da sette nobili fiorentini nel 1384. Gli autori dei testi esaminati sono Lionardo Frescobaldi e Giorgio Gucci, rivalutati in quanto autori «del buon secolo» nell'ultima fase (intorno alla metà del XIX secolo) della plurisecolare attività lessicografica dell'Accademia della Crusca.

Di entrambi si possiede una *editio princeps* ottocentesca: quella di Frescobaldi venne approntata a Roma da Guglielmo Manzi nel 1818, quella di Gucci a Firenze da Carlo Gargioli nel 1862 (in questa stessa stampa è contenuto il testo di Frescobaldi allestito secondo ragioni filologiche diverse rispetto all'edizione del 1818). Nella prima parte del contributo si ricostruiscono le modalità di allestimento di queste edizioni, e la considerazione che di esse si aveva negli ambienti intellettuali fiorentini intorno alla metà del XIX secolo.

Nella seconda parte del saggio sono condotte osservazioni puntuali sulla presenza di questi resoconti nelle opere lessicografiche ottocentesche, in particolare nella V Crusca, grazie agli spogli di Gaetano Milanese e Giovanni Tortoli, e nel Tommaseo-Bellini (con osservazioni relative anche ai vocabolari di Fanfani e Manuzzi). L'esame conferma da un lato il fortissimo legame esistente all'epoca tra lavoro filologico e lessicografico, dall'altro la prassi che caratterizza a lungo la lessicografia italiana, cioè la tendenza a operare *giunte* ai vocabolari precedenti, in particolare alla IV Crusca.

This essay examines the philological and editing history and the nineteenth century lexicographical receptions of two accounts of a journey to the Holy Land made by seven Florentine noble men in 1384. The authors of the texts examined are Lionardo Frescobaldi and Giorgio Gucci, revalued as authors «del

buon secolo» in the last phase (roughly half way through the nineteenth century) of the centuries-old lexicographical activity of the Accademia della Crusca.

A nineteenth century *editio princeps* of both texts is available: Frescobaldi's edition was prepared by Guglielmo Manzi in 1818, while Gucci's was made in Florence by Carlo Gargioli in 1862 (this printed edition also includes the text by Frescobaldi, prepared on the basis of different philological criteria from those used for the edition of 1818). The first part of the article retraces the way in which these editions were prepared, and the attention given to these works in the intellectual environment of mid-nineteenth century Florence.

The second half of the essay gives a detailed analysis of these accounts in nineteenth century lexicographical works, especially in the fifth edition of the Vocabulary of the Accademia della Crusca, thanks to the perusals made by Gaetano Milanese and Giovanni Tortoli, and in the dictionary by Tommaseo and Bellini (examining also the dictionaries written by Fanfani and Manuzzi). The analysis confirms on the one hand the strong bond existing at the time between philological and lexicographical work, on the other, the practice that has long been a characteristic of Italian lexicography, in other words the tendency to add supplements to previous dictionaries, especially in the fourth edition of the Vocabulary of the Accademia della Crusca.

YORICK GOMEZ GANE, «Sballare»: approfondimenti storico-linguistici e lemmatizzazione

Il verbo *sballare* presenta diversi significati, alcuni dei quali si ramificano, lungo un arco cronologico molto ampio (dal XV secolo ai nostri giorni), in numerose accezioni. I principali vocabolari storici ed etimologici hanno adottato soluzioni diverse circa la suddivisione dei significati fondamentali (con conseguente divergenza nella registrazione dei lemmi omonimici) e in rapporto alla derivazione delle accezioni dall'uno o dall'altro dei significati. In questa sede, passando sistematicamente in rassegna significati e accezioni del verbo si propone, in ragione di nuove osservazioni etimologiche e col supporto di nuovi dati storico-linguistici, una specifica lemmatizzazione dei significati principali (*sballare*¹ 'togliere dall'imballo', *sballare*² 'in alcuni giochi di carte, oltrepassare il numero dei punti stabiliti, perdendo la posta', *sballare*³ 'portare su di giri (un motore)' e *sballare*⁴ 'interrompere (un ballo)'), registrando e analizzando le relative accezioni derivate, con particolare attenzione a origine ed evoluzione semantica delle singole forme.

The verb *sballare* has many meanings, some of which ramify, over a very wide time-span (from the fifteenth century to the present day), in numerous instances. The most important historical and etymological dictionaries have found different solutions for distinguishing the most important meanings (therefore

there is a difference in recording the homonymic lemma), and, concerning the derivation of the instances from one or other of the meanings. In these dictionaries, analysing in depth meanings and instances of the verb, due to new etymological considerations and on the basis of new historical and linguistic data, a detailed lemmatization is provided for the main meanings *sballare*¹ 'togliere dall'imballo', *sballare*² 'in alcuni giochi di carte, oltrepassare il numero dei punti stabiliti, perdendo la posta', *sballare*³ 'portare su di giri (un motore)' e *sballare* 'interrompere (un ballo)'. The related instances derived are recorded and analysed, paying special attention to the origin and semantic evolution of each form.

GIUSEPPE POLIMENI, Carlo Gambini, il dialetto pavese, la questione della lingua in Italia

Il saggio si propone di avvicinare la figura e l'opera di Carlo Gambini (1799-1886), uomo di diritto e magistrato, attento fin dagli anni giovanili alla realtà del dialetto e ai problemi connessi al lessico dell'italiano in una fase di transizione (dal mondo della Restaurazione all'Unità d'Italia), significativa sia sul piano politico, sia su quello sociale e culturale.

In questo contributo, che offre alcuni sondaggi in vista di una più ampia trattazione monografica del personaggio e del suo pensiero linguistico, si prendono in considerazione le dichiarazioni teoriche che accompagnano la sua attività di "compilatore" del dizionario del dialetto pavese (nelle tre edizioni del 1829, 1850, 1879) e si porta quindi l'attenzione sulla sensibile analisi del linguaggio dell'amministrazione pubblica, per dedicare alcune pagine finali alla presa di posizione netta (a favore di Pietro Fanfani e di Luigi Gelmetti) con cui, a partire dal 1878, Gambini entra nel dibattito aperto nel 1868 dalla *Relazione* manzoniana.

This contribution draws attention to the personality and works of Carlo Gambini (1799-1886), lawyer and judge, who since early youth made a study of dialect and problems connected with the Italian lexicon in a period of transition (from the Restoration to the Unification of Italy), that was important from both a political and a socio-cultural point of view.

This essay provides some preliminary enquiries with a view to a deeper monographic study of the person and his linguistic thoughts; the theoretical principles that accompany Gambini's work of compiling a dictionary of the dialect of Pavia (with three editions, in 1829, 1850 and 1879) are examined and attention is focused on his sensitive analysis of the language of public administration. The last pages are devoted to the clear stand (in favour of Pietro Fanfani and Luigi Gelmetti) that Gambini took in the debate that was initiated in 1878 by Manzoni's *Relazione*.

STEFANO CALONACI, Tra storia, educazione popolare e filologia: la formazione di Pietro Fanfani polemistà e lessicografo

Nel saggio si cerca di ricostruire la formazione di Pietro Fanfani, considerata sotto il profilo dell'attività di lessicografo e della visione che egli ebbe della trasmissione culturale come momento educativo. Si ripercorrono quindi le esperienze culturali, pedagogiche e gli incontri del giovane Fanfani, nonché l'acquisizione di una pratica di lettura paleografica esercitata sui manoscritti di archivi e biblioteche che sarebbe divenuta viatico di una personale riflessione sulla lingua. Nella biografia intellettuale del Fanfani viene così individuata l'incidenza di figure dai profili molto differenziati e allo stesso tempo tipici del contesto sociale toscano di inizio Ottocento: il prete di paese, la maestra beghina, il sacerdote colto e letterato. Ma anche il contatto con letterati provenienti da altre realtà culturali, come Giuseppe Silvestri. Assieme ai libri e alla lettura, il loro magistero e la loro frequentazione precedettero in Fanfani l'esperienza educativa più istituzionalizzata del Seminario diocesano e della Scuola ospedaliera. Ne emerge, assieme al profilo di Fanfani, un vario panorama di uomini e istituti della trasmissione del sapere attivi nel quadro dell'organizzazione sociale ed economica del territorio, animando i tempi della vita secondo un patrimonio linguistico d'uso comune, sullo sfondo del più ampio processo di unificazione linguistica e nazionale.

This essay aims at retracing Pietro Fanfani's education, considered from the viewpoint of his lexicographical activity and of his vision of the transmission of knowledge as an educational opportunity. The cultural and pedagogic experiences and encounters of the young Fanfani are thus retraced, as well as his acquisition of a paleographical reading technique used for manuscripts in archives and libraries that would later become his viaticum for his own reflections on language. The intellectual biography of Fanfani reveals the influence of figures who are very diverse and yet characteristic of the social context in early nineteenth century Tuscany: the village priest, the Beguine teacher, the cultivated and literate priest. It includes also the contact with intellectuals from other cultural environments, such as Giuseppe Silvestri. Alongside the books and reading, their teaching and their frequentation preceded Fanfani's more institutional education at the diocesan Seminary and the hospital School. Together with this profile of Fanfani, a varied panorama emerges of the men and institutions for transmitting knowledge that were active in the social and economic organisation of the territory. This brings life to the times according to a common linguistic heritage, against a background of the wider process of linguistic and national unification.

ANTONIO VINCIGUERRA, *Le inedite aggiunte e correzioni di Emmanuele Rocco ai vocabolari italiani: descrizione dei materiali e sondaggi lessicali*

L'articolo presenta e descrive i nove volumi manoscritti di Emmanuele Rocco, conservati nell'Archivio storico dell'Accademia della Crusca, contenenti circa ventimila inedite aggiunte, correzioni e osservazioni varie ai vocabolari italiani sette-ottocenteschi (dalla IV Crusca alle prime voci del Tommaseo-Bellini), le quali furono raccolte dal filologo napoletano nell'arco di un trentennio che va dal periodo della sua collaborazione al Tramater fino ai primi anni postunitari. Tali volumi costituiscono ancora oggi un documento di notevole interesse lessicografico e storico-linguistico, non solo perché furono in parte utilizzati dagli accademici compilatori della quinta impressione del Vocabolario della Crusca, ma anche, e soprattutto, perché registrano numerosi termini e usi (ricavati dallo spoglio di opere letterarie come pure di fonti d'altro tipo, quali periodici e testi di carattere specialistico) che risultano assenti o scarsamente documentati nei dizionari generali dell'italiano otto-novecenteschi. Nel saggio sono stati trascelti e trattati, per mezzo di specifici commenti, in particolare gli usi neologici, i termini scientifici e quelli settoriali registrati nei volumi di Rocco.

This article presents and describes the nine hand-written volumes by Emmanuele Rocco, kept in the historical Archive of the Accademia della Crusca, which include roughly twenty thousand, unpublished additions, corrections and various observations concerning eighteenth and nineteenth century Italian dictionaries (from the fourth edition of the *Vocabolario degli Accademici della Crusca* to the first entries of the dictionary by Tommaseo and Bellini), which were collected by the Neapolitan philologist over thirty years, from the time of his collaboration for Tramater until the first years after the unification of Italy. These volumes still represent nowadays a document of great lexicographical and historical-linguistic interest, not just because they were partly used by the academicians for drafting the fifth edition of the *Vocabolario* of the Accademici della Crusca, but also, and especially, because they record many words and uses (obtained from the perusal of literary works and also of other sources, such as periodicals and specialised texts) that are absent or with rare documentation in the general eighteenth and nineteenth century Italian dictionaries. In this essay, some neologisms, scientific terms and sectorial ones registered in Rocco's volumes have been selected and analysed, with specific comments.

FRANCESCA MALAGNINI - ANNA RINALDIN, Cronologia esplicita e nuovi dati redazionali per il «Dizionario della lingua italiana» di Niccolò Tommaseo e Bernardo Bellini: l'esemplare in dispense

A partire dall'unico esemplare conservato in dispense del *Dizionario della lingua italiana* di Niccolò Tommaseo e Bernardo Bellini e ospitato presso la biblioteca del Centro interuniversitario di studi veneti (CISVe) dell'Università Ca' Foscari di Venezia, il lavoro che qui si presenta offre la cronologia esplicita e nuovi dati redazionali dell'opera stampata tra il 1861 e il 1879. L'esemplare in dispense (di cui si dà anche una descrizione materiale) consente di precisare la cronologia delle uscite, e quindi delle voci comprese nelle singole dispense. Emergono inoltre alcune notizie sui rapporti con l'editore Pomba e con la rete dei collaboratori all'impresa; infine, da alcune recensioni riportate sulla copertina delle dispense si ricavano indicazioni sull'accoglienza dell'opera durante la sua composizione.

The only extant witness of the *Dizionario della lingua italiana* by Niccolò Tommaseo and Bernardo Bellini – preserved in its original form of single bundles at the library of the Centro interuniversitario di studi veneti (CISVe) at the University Ca' Foscari in Venice – is a relevant one to understand the stages of drafting of the dictionary between 1861 and 1879. The essay provides a description of the witness and a precise chronology of printed bundles, so that for each single entry of the work can be given a year of publication. It is also possible, by means of the witness, to enlighten the arrangements with the publisher, Giuseppe Pomba, and with Tommaseo's network of contributors. Moreover, on bundles' covers were copied some reviews of the time, which testify the audience reception of the dictionary during its publication.

MIRKO VOLPI, Mantegazza onomaturgo. Note lessicali su «L'anno 3000. Sogno»

Il saggio prende in esame il lessico del romanzo fantascientifico *L'Anno 3000. Sogno* (1897) di Paolo Mantegazza, rilevante in particolare per ciò che concerne il linguaggio della divulgazione scientifica nell'Italia del secondo Ottocento. Si analizzano, nello specifico, le modalità attraverso cui l'autore crea nuove parole tecniche d'ambito medico (perlopiù attraverso composti neoclassici), che in taluni casi si affermeranno, e rimarranno, anche nella lingua d'uso; e la notevole presenza di tecnicismi di svariati settori (dalla psicologia alla botanica, dalla meccanica alla biologia, ecc.), prime occorrenze o di recentissimo ingresso in italiano (tra cui anche interessanti “darwinismi”), che Mantegazza, qui come in altri suoi testi (vedi il *Secolo nevrosico*, 1887, di cui pure si occupa questo contributo), sa tempestivamente intercettare e riutilizzare nelle sue opere divulgative.

This essay examines the lexicon of the sci-fi novel *L'anno 3000. Sogno* (1897) by Paolo Mantegazza, which is especially relevant for the language of scientific diffusion in late nineteenth century Italy. In particular, an analysis is made of the means by which the author creates new technical words from the field of medicine (mostly by means of neoclassical composite words): in some cases these words were to become part of the current language and remain there. The widespread occurrence of technical terms from different fields (from psychology to botany, from mechanics to biology, etc.) is also noted. Some are early examples while others have recently become part of the Italian language (including some interesting examples from Darwin), that Mantegazza, here as in other works (for example *Secolo nevrosico*, 1887, also analysed in this essay), knew how to capture and re-use in his most popular works.

PIERO FIORELLI, Cent'anni d'ortopedia toponomastica

Nell'ottobre 1916 viene pubblicato dal Touring Club Italiano l'*Indice generale della Carta d'Italia del T.C.I. alla scala 1: 250.000, compilato sotto la direzione di L[uigi] V[ittorio] Bertarelli*. Questo articolo vorrebbe servire a fare apprezzare quanto è giusto l'utilità e il peso che quella *Carta* e soprattutto quell'*Indice* hanno avuto nell'accertare e far conoscere il più alto numero possibile di toponimi italiani, fissandone una volta per tutte l'esatta forma grafica (e insieme, ridotta all'essenziale, fonetica) così da servire da obbligato punto di riferimento per la cartografia, per gl'indici e i repertori, per gli studi di toponomastica e anche di lessicografia.

In October 1916 the Touring Club Italiano published the *Indice generale della Carta d'Italia del T.C.I. alla scala 1: 250.000, compilato sotto la direzione di L[uigi] V[ittorio] Bertarelli*. This essay aims at showing the importance and usefulness that the Atlas and especially the *Indice* had in ascertaining and making accessible the largest number of Italian toponyms, establishing once and for all their precise written form (as well as the essential phonetic form) in order to provide a fundamental reference point for cartography, indexes and repertories, toponymic and lexicographical studies.

CHIARA COLUCCIA - MARIA VITTORIA DELL'ANNA, Lingua italiana e ambiente. Note sul lessico dell'ecologia

La ricerca si offre come contributo alla discussione su lingua e lessico dell'ecologia nel nostro Paese, a cominciare dall'indagine sulla presenza e sul trattamento del lessico collegato ad 'ambiente' ed 'ecologia' nei dizionari dell'uso della lingua italiana. I dati offerti dai dizionari offrono lo spunto per osserva-

zioni lessicografiche e lessicologiche concentrate intorno ai seguenti aspetti: problemi relativi alla costituzione del lemmario ecologico, legata in termini qualitativi e quantitativi agli apporti lessicali dalle discipline che concorrono storicamente alla formazione e allo sviluppo della tematica ecologico-ambientale; statuto settoriale del lessico marcato come ecologico; tendenze diffuse nella formazione delle parole; apporti da altre lingue.

This research makes a contribution to the debate on language and the lexicon of ecology in Italy, starting with an enquiry into the presence and the treatment of the lexicon linked with 'environment' and 'ecology' in Italian dictionaries in current use. The data provided by the dictionaries provide an input for lexicographical and lexical comments that focus on the following aspects: problems involved in the formation of a list of ecological words, linked in quality and quantity to the lexical contributions from the subjects that have historically contributed to the formation and development of ecological and environmental issues; the specific status of the ecological lexicon; widespread tendencies in the formation of words; contributions from other languages.

EMANUELE VENTURA, Note sul lessico ciclistico contemporaneo: fra gergo e lingua quotidiana, fra tradizione e innovazione

L'articolo, fondandosi su un ampio glossario, analizza il lessico contemporaneo del ciclismo, caratterizzato da una particolare commistione di voci tradizionali e voci di più recente diffusione, tra le quali si osservano alcuni anglicismi, ma soprattutto molti termini derivati dalla lingua comune e andati incontro a un processo di risemantizzazione: negli ultimi decenni, in particolare, hanno assunto una posizione sempre più rilevante molte parole ed espressioni provenienti dal gergo degli atleti (professionisti e amatori) e successivamente accolte in pianta stabile anche nella lingua giornalistica nazionale.

This article, based on an extensive glossary, analyses the contemporary lexicon of cycling, characterized by a peculiar mix of traditional words, as well as more recent ones. Among these words some Anglicisms are observed but, above all, many terms derived from the common language that went through a process of re-semanticisation: in the last decades, in particular, many words and expressions used by both professional and amateur athletes as jargon have taken on an increasingly important position and, subsequently, they have also been integrated on a stable basis in the language of Italian journalism.

PAOLO D'ACHILLE, «A te l'estremo addio»? Il problema dell'ultima attestazione nella linguistica e nella lessicografia italiana

Sulle prime attestazioni delle parole italiane gli studi sono in pieno sviluppo, con continue retrodatazioni. Invece la lessicografia e la lessicologia hanno dedicato scarsa attenzione alle ultime attestazioni di parole e forme uscite dall'uso: di fatto l'unico punto di riferimento è costituito dal *Grande dizionario della lingua italiana* del Battaglia (*GDLI*), e dagli ultimi esempi che riporta sotto le singole entrate. Lo studio vuole costituire una prima riflessione su questo tema, presentando una serie di parole (sia grammaticali, sia semanticamente piene), di cui esistono attestazioni posteriori a quelle segnalate dal *GDLI*. Alla fine si propone qualche considerazione generale sul concetto e il significato dell'ultima attestazione, nonché sulla difficoltà di individuarla, nel quadro del complesso rapporto tra l'italiano della tradizione e la lingua contemporanea.

The investigation of early occurrences of Italian words is in full swing, constantly providing the opportunity to backdate their first attestations. Lexicography and lexicology have, however, paid less attention to the last attestation of words and forms which have fallen out of use: on this point the only data available are in practice those provided by Battaglia's *Grande dizionario della lingua italiana* (*GDLI*), specifically the last references provided there under the individual entries.

This study considers some of the problematic aspects of this topic, exemplified by reference to a number of word (both grammatical and lexical) for which occurrences later than those indicated by *GDLI* are attested. Some general considerations on the notion of last attestation and its valence are furthermore provided, as well as on the difficulties involved in identifying the last attestation of a word in the context of the complex relationship between literary Italian and contemporary usage.

MARCO BIFFI, ALICE FERRARI, Progettare e realizzare un «corpus» dell'italiano nella rete: il caso del «CoLIWeb»

All'interno delle varie fasi di realizzazione del *VoDIM* (*Vocabolario dinamico dell'italiano moderno*), che discende dalla ricerca effettuata per Prin specifici nel 2012 e nel 2015, si colloca la creazione di un *corpus* diacronico comprendente la lingua della scienza, dell'arte, della cucina, della politica, delle canzoni e delle opere liriche, della letteratura e della paraletteratura, dell'informazione, delle istituzioni. Per aumentarne il grado di rappresentatività è parso opportuno creare, in aggiunta, un *corpus* bilanciato delle dimensioni di 2 miliardi di parole riferito all'italiano più recente sulla base del web. Da qui è nato il progetto di realizzazione di un *corpus* con queste caratteristiche attra-

verso strumenti informatici che consentono di prelevare e catalogare ingenti quantità di testi dalla rete. L'articolo proposto si sofferma in particolare sul metodo di impostazione del programma di prelievo, basato sulle "etichette" del *Grande dizionario italiano dell'uso* di De Mauro, e sulla descrizione del primo prototipo di banca dati interrogabile

The various stages in the compilation of the *Vocabolario dinamico dell'italiano moderno (VoDIM)*, which developed from research carried out for national projects of 2012 and 2015, include the creation of a diachronic *corpus* of the language of science, art, cooking, politics, songs and operas, literature and para-literature, information, institutions. To increase the degree in which the *corpus* was representative, it seemed appropriate to create, in addition, a balanced corpus of the size of 2 billion words with reference to the most recent web-based Italian. This led to the project of creating a corpus with these characteristics by using the tools of information technology that allow the extraction and cataloguing of large quantities of texts from the web. This article concentrates in particular on the method of designing the programme for extracting texts, based on the "labels" of De Mauro's *Grande dizionario italiano dell'uso* and on the description of the first prototype of searchable data-base.

(traduzioni in inglese a cura di Matteo Gaja)