Computational Lexicography and Carlo Emilio Gadda, *Principe dell'Analisi e Duca della Buona Cognizione*

Maria Luigia Ceccotti, Manuela Sassi

ILC-CNR
Via Moruzzi, 1 - 56100 Pisa
[luisia, manuela.sassi]@ilc.cnr.it

**Abstract**

The aim of this study is to stimulate interest in searching the Electronic Archive of Carlo Emilio Gadda’s Works and the web site dedicated to this important Italian author in order to receive and implement suggestions, and to create new lexical resources for students, translators and readers. The article starts with a brief description of the editorial features of this Archive, that runs under the 2004 version of the Textual DataBase (DBT), and continues with the presentation of the web site, to be intended as a cultural laboratory containing texts by Gadda, bibliographical data, links and special lexicographic resources developed in three ways: by applying simple DBT functions (Hapax Legomena List, Index Locorum and Concordances), by submitting results to the semi-automatic lemmatization process (Concordances lemmatized), and by using multiple Archive outputs to be transformed into new DBT archives.

1. **Background**

The project for the creation of the “Electronic Archive of Carlo Emilio Gadda’s Works” was started in 1994, when publisher Garzanti made their published text materials (Gadda, 1988-93), already prepared for photo-composition and for study purposes only, available to the Istituto di Linguistica Computazionale (ILC, Institute of Computational Linguistics).

1.1. **The first version of the Corpus**

The Corpus was created starting from these texts (33), appropriately encoded under DBT¹ format and by applying a mark-up to point out various phenomena (italic characters, capital words of proper nouns or decided by the author, hyphens, dates, numbers, formulas, abbreviations, acronyms, pictures, author’s and editor’s footnotes, dialect and foreign words, dialogue, poetic language). The first version of the Corpus, running in DBT 3.0, was presented on November 14, 1997 (the 104th anniversary of Gadda’s birthday) at the CNR (National Research Centre) in Rome.

1.2. **The first results**

In 1997 we became the first users of this database, which is the only electronic archive existing to contain all the works of a contemporary Italian author. While the community of Gadda scholars is starting to demand something that cannot be distributed but is available for consultation at ILC – concordances -, we have started to:
- produce lexicographic tool standards by means of an automatic processing system;
- apply innovative methodologies onto sample subsystems.

2. **The CEG Project**

The CEG (Carlo Emilio Gadda) Project was created in 1999 with two objectives:
- to create a cultural laboratory containing texts by Gadda, bibliographical data, links and special lexicographical resources.

2.1. **The Textual DataBase (DBT)**

We will now briefly describe the peculiar features of the DBT 2004, the most authoritative heir and witness of ILC’s thirty-year experience in the sector of computational lexicography. In addition to the functions existing since the first version of the programme, the most recent version of this textual database also allows for searches by lemma, by synonym, by complex family of forms and/or lemmas, by specialised words, and by descriptors. However, the DBT2004’s potential for application to texts written in a given language is based on the “results” provided by two lexicographic instruments, the dictionary and the Thesaurus, for the first three types of search, and on entering specific codes into the texts examined, in the last two types of search. As regards the consultation of Gadda’s archive by lemma, the dictionary and Thesaurus that are presently referred to are those of general use. However, we deem it necessary to design a project for the creation of both a Gadda dictionary and a Gadda thesaurus.

Obviously, for an author like Gadda, whose lexicon is rich of variants, archaisms, neologisms, dialect terms, etc., the results provided by the database after a search by lemma or by synonym are not always exhaustive.

2.1.1. **Search by lemma**

If we perform a search with the word *patire*, the first result we obtain is a list of the forms retrieved in the texts, with an indication of the number of texts of the corpus in which they appear:

```
patente (10) patenti (5)
pati (10) [Apati (1)
pati (1) patir (6)
pair (1) patire (12)
|Apater (1) patiro (1)
patirone (1) patisce (7)
patisco (1) patiscono (3)
patisse (2) patita (9)
patite (7) patiti (4)
|Apatis (1) patito (14)
|Apatic (1) pativa (8)
patiano (3) pativo (2)
```

¹ CNR Patent by Eugenio Picchi, chief-researcher of the ILC.
2.1.2. Search by synonym

From the general use Thesaurus, searching with the synonym function, will deliver the adjectival forms labelled as synonyms of the adjective *utile*:

- efficace (11)
- efficaci (5)
- efficacissime (2)
- efficacissimo (1)
- validamente (4)
- valide (2)
- validissime (1)
- validissimo (1)
- vantaggiosa (2)
- vantaggiosi (1)
- opportuna (12)
- opportuno (8)
- proficusa (5)
- proficuamente (2)
- salutare (2)
- giovevole (2)

2.1.3 Search by complex family (forms and/or lemmas)

When more complex searches by groups of words / lemmas are required, the system will give the contexts retrieved by the search for *patire and fame*:

1) *Hapax Legomena Inverse Index* (Ceccotti, 1998):
   - alpha / omega (abeto - capataz)
   - beta / gamma (aricordamo - novantuno)
   - delta / epsilon (aio - lapislazzulo)
   - epsilon / zeta (ciciàk - serrucho)
   - zeta / epsilon (crepi - complimentuzzi)
   - epsilon / gamma (esofàghi - vvoi)
   - gamma / delta (ciàaf - arciduchi)
   - delta / epsilon (ère - elegantuzze)
   - epsilon / delta (recane - scolopendre)
   - delta / eta (sub - frustume)
   - eta / delta (senna - sapienzuzza)
   - delta / zeta (sub - frustume)
   - zeta / delta (senna - sapienzuzza)
   - delta / zeta (sub - frustume)
   - zeta / delta (senna - sapienzuzza)
   - delta / zeta (sub - frustume)
   - zeta / delta (senna - sapienzuzza)

2) Bestie pazze! per cui ho *patito il freddo* e la fame per le pere, non sa neanche lui cosa dice - RR1-CD.1.III.152.p.0612.5

3) i brillanti e che loro hanno *patito* il freddo e la fame per le pere, non sa neanche lui cosa dice - RR1-CD.1.III.652.p.0636.37

4) Bestie pazze! per cui ho *patito la fame*, da bimbo, la fame! Cinquecento pesos! Cinquecento: di munificenza - RR1-CD.1.III.652.p.0636.37

5) i brillanti e che loro hanno *patito* il freddo e la fame per le pere, non sa neanche lui cosa dice - RR2-AG.7.110.p.0719.37

6) Bestie pazze! per cui ho *patito la fame*, da bimbo, la fame! Cinquecento pesos! Cinquecento: di munificenza - RR2-AG.7.941.p.0741.12

7) i brillanti e che loro hanno *patito* il freddo e la fame per le pere, non sa neanche lui cosa dice - RR2-AG.7.110.p.0719.37

8) un articolo rabberciato da un precedente, dovremo dire che *pativa la fame*? Il gioco è stato bello: e gioco sia - SGF1-SD.27.147.p.0829.24

9) pazienza, senza alcuna inspirazione. Gran debolezza fisica: patii molto la fame, come il solito: a cena un mestolo di - SGF2-GGP.D17.596.p.0673.19

10) altro di viveri di riserva. - Sebbene prevedessi la fame che avremmo *patita*, preferir le vesti ai viveri. Feci male, - SGF2-GGP.M.1615.p.0739.7

2.2. The web site

In 1999, when we had become more familiar with computer graphics and HTML or XML formats, we began to plan the creation of a web site on the works written by Gadda. Although we had not received an *ad hoc* funding, at the beginning of 2000 we started to create the CEG web site exclusively based on our know-how. The web site advertised the history of the project and made our first paper publications available online.

At present, the web site’s home page presents a menu that the user can use to perform searches by:

- **MAP** [MAPP] - the site map, which is regularly updated
- **PRESENTATION** [PRESENTAZIONE] - the presentation of the project and its objectives
- **REFERENCES** [BIBLIOGRAFIA] - a brief secondary bibliography
- **CHRONICLE** [CRONOGRAFIA] - here we store important dates (the start of a publishing project, data entry, etc.)
- **LINKS** - links to significant web sites or any other reference to Gadda in the web
- **REPORT** - our HTML introductions to Gadda-related data collections and some monographs
- **LOOK UP [CONSULTA]** - collections of Gadda-related data, both in XML and HTML
- **CONCORDANCES** [CONCORDANZE] - the traditional concordances produced up to today, in PDF
- **INSTRUMENTS** [STRUMENTI] - lexicographic instruments obtained by processing DBT function outputs
- **DEMO** - a demo of the Corpus, which can be consulted online with a DBT-Web interface built in compliance with copyright restrictions
- **SHORTVIEW** - slides used to present the project at several conferences. Furthermore, a NEWS section can be directly reached from the home page where the last works completed or in progress are shown.

3. Linguistic Resources

The linguistic resources are created based on our thirty-year experience in textual data processing and on the suggestions obtained either from readings or from researchers, teachers and students who have looked at new technologies for classical studies through the intermediation of ILC.

3.1. Linguistic resources available

The first group of the listed resources has been produced in both electronic and hard copy format, while the second only exists in the electronic format.

1) Hapax Legomena Inverse Index (Ceccotti, 1998):
3.2. Linguistic resources in progress

While in the first phase of our database usage activity we have produced lexicographic resources mainly by applying simple system functions, the method we are prevalently using at present consists in processing the results obtained with different instruments.

3.2.1. Contrastive concordances

The experience made with the construction of a comparison between the two versions of Quer pasticciocecco brutto de via Merulana - the first (QPL) being issued in installments in the years 1946-47 in the Letteratura magazine and the second (QP) published by Garzanti in 1957 - was very useful to identify a method to perform virtually automatic comparisons between multiple text versions. These two texts stored in the archive have been isolated and used to create a sub-corpus. The DBT-Corpus table frequency function has been used to generate a table containing a first comparison between words. In the subsequent step, the forms found in both texts were ignored, while only those found in one of the two texts were taken into consideration, namely those of the second version. These data have been re-entered into the search system to extract the relevant contexts, which have been matched by highlighting the differences between them, as shown in Table 2.

Table 1: QP and QPL matched contexts

| QP-17 | QPL-283 |
| --- | --- |
| può stà ssicure ch'è nu guai: quacche glinommero,de sherrêtâ ... | può stà ssicure ch'è ca’ gguaio ..., quacche glinommero ... de sherrêtâ ... |
| QP-17 | QPL-283 |
| soffiate addosso a molinello (come i sedici venti della rosa dei venti) | soffiate addosso a molinello (come i 16 venti della rosa dei venti) |
| QP-17 | QPL-283 |
| «ch'i femmene se retroveno addó n'i vuò truòvà». che 'e femmene se retroveno addo’ n’i vuò’ trovà. | |
| QP-17 | QPL-283 |
| qualche prete più edoto dei molti danni del secolo, alcuni subalterni, certi uscieri, i superiori, sostenevano che | qualche prete suo conoscente, gli uscieri, i superiori, sostenevano che |
| QP-17 | QPL-283 |
| come non altre ad acciecurre gli sprovveduti, gli ignari. | ma servono a gettâr la polvere negli occhi alle genti. |

Table 2: Most frequent iterations
3.3. The future of linguistic resources

Many more data have been collected and are currently available, and we hope we will be soon able to add the resources produced by the application of the tagger to the whole archive in our web site. In particular, we are extracting adjectives, which are numerous and sometimes neologisms created by Gadda himself, and represent one of the most interesting testimonies of the lexicon of this important author of the Italian twentieth-century literature.

In addition to the objective of recording the lexical universe of the author, we also aim at creating supports that may be useful, for instance, for translators and teachers. All the co-occurring adjectives for the substantive prosa have been extracted:

- acre
- avanzata
- brillante
- burocratica
- commovente
- curiale
- distratta
- elegantesima
- fluent
- generica
- icastica
- inarrivabile
- insuperata
- magnifica
- meravigliosa
- nervosa
- pacata
- peggiore
- piscatoria
- polligiosa
- preferibile
- sapida
- secca
- sorvegliata
- tersa
- vigorosa

We are presently planning to create conceptual structures and apply them to Gadda’s works through the DBT. But some explanation is necessary in this regard. The DBT statistic co-occurrence function allows us both to capture some explanation is necessary in this regard. The DBT statistic co-occurrence function allows us both to capture some explanation is necessary in this regard. The DBT statistic co-occurrence function allows us both to capture some explanation is necessary in this regard. The DBT statistic co-occurrence function allows us both to capture some explanation is necessary in this regard. The DBT statistic co-occurrence function allows us both to capture some explanation is necessary in this regard. The DBT statistic co-occurrence function allows us both to capture some explanation is necessary in this regard. The DBT statistic co-occurrence function allows us both to capture some explanation is necessary in this regard. The DBT statistic co-occurrence function allows us both to capture some explanation is necessary in this regard. The DBT statistic co-occurrence function allows us both to capture some explanation is necessary in this regard. The DBT statistic co-occurrence function allows us both to capture some explanation is necessary in this regard. The DBT statistic co-occurrence function allows us both to capture some explanation is necessary in this regard. The DBT statistic co-occurrence function allows us both to capture some explanation is necessary in this regard. The DBT statistic co-occurrence function allows us both to capture some explanation is necessary in this regard. The DBT statistic co-occurrence function allows us both to capture some explanation is necessary in this regard. The DBT statistic co-occurrence function allows us both to capture some explanation is necessary in this regard. The DBT statistic co-occurrence function allows us both to capture some explanation is necessary in this regard. The DBT statistic co-occurrence function allows us both to capture some explanation is necessary in this regard. The DBT statistic co-occurrence function allows us both to capture some explanation is necessary in this regard. The DBT statistic co-occurrence function allows us both to capture some explanation is necessary in this regard. The DBT statistic co-occurrence function allows us both to capture some explanation is necessary in this regard. The DBT statistic co-occurrence function allows us both to capture some explanation is necessary in this regard. The DBT statistic co-occurrence function allows us both to capture some explanation is necessary in this regard. The DBT statistic co-occurrence function allows us both to capture some explanation is necessary in this regard. The DBT statistic co-occurrence function allows us both to capture some explanation is necessary in this regard. The DBT statistic co-occurrence function allows us both to capture some explanation is necessary in this regard. The DBT statistic co-

4. Conclusion

We would like to conclude by quoting Gadda himself, in a passage that seems to us to be appropriate to comment upon the presentation of this contribution of ours to a conference where analysis and good cognition are important reference parameters: “Com’è noto sono stato recentemente insignito del titolo nobiliare ereditario di Principe dell’Analisi e di Duca della Buona Cognizione per motivi che tutti conoscono e con un rituale che sarà oggetto d’un particolare e distinto referto. L’impresa del Principe dell’Analisi comporta il motto brunianno (della Cena delle ceneri): «Umbra profonda sumus» ed è questa secondo alcuni la più dolorosa conclusione alla quale possa pervenire un indagatore. Essa però non deve interpretarsi in senso scettico: una constatazione d’ombra non chiude la dolorosa fatica del minatore. Così come i vincoli o legami non impidirono a Michelangiolo di esprimere il suo Schiavo legato.” (da La casa, Racconti incompiuti, RR2, p.1127; Gadda, 1989)

“As is well-known, I have recently been conferred the hereditary noble title of Prince of Analysis and Duke of Good Cognition for reasons that everybody knows and with a ritual that will be the subject of a special and distinct report. The deeds of the Prince of Analysis involve Bruno’s motto (in the Cena delle ceneri): «Umbra profonda sumus» and this is, according to some, the most painful conclusion that could be reached by an investigator. But it should not be interpreted as a sceptical statement: the ascertainment of a shadow will not close the painful fatigue of the miner. Similarly, bonds or ties did not prevent Michelangiolo to express his tied Slave.”

References

Ceccotti, M.L. & Sassi, M. (1998). Apax in Gadda – Un Indice Inverso, ILC-CNR, S.T.A.R., Pisa.

Ceccotti, M.L. & Sassi, M. (1999a). Forme accentate in Gadda. Istituti Editoriali Poligrafici Internazionali, Pisa.

Ceccotti, M.L. & Sassi, M. (1999b). Alla ricerca dei termini gaddiani. Un dataset, ILC-CNR, S.T.A.R., Pisa.

Ceccotti, M.L. & Sassi, M. (2002a). Sistema. In A Pocket Gadda Encyclopedia. Edited by Pedriali F.G. EJGS Supplement no. 1, ISSN 1476-9859.

Ceccotti, M.L. & Sassi, M. (2002b). La cultura latina in Gadda. ILC-CNR, S.T.A.R., Pisa.

Ceccotti, M.L. & Sassi, M. (2003) L’Archivio Elettronico delle Opere di Carlo Emilio Gadda. Da redattori a fruitori di un data base testuale, in Linguistica Computazionale, Special Issue (Zampolli A., Calzolari N., Cignoni L. Editors), Istituti Editoriali Poligrafici Internazionali, Pisa-Roma, pp.221--250.

Gadda, C.E. (1988). Romanzi e Racconti I, Collana I Libri della Spiga, Garzanti, Milano.

Gadda, C.E. (1989). Romanzi e Racconti II, Collana I Libri della Spiga, Garzanti, Milano.

Gadda, C.E. (1991). Saggi Giornali Favole I, Collana I Libri della Spiga, Garzanti, Milano.

Gadda, C.E. (1992). Saggi Giornali Favole II, Collana I Libri della Spiga, Garzanti, Milano.

Gadda, C.E. (1993). Roma e Racconti, Collana I Libri della Spiga, Garzanti, Milano.

Gadda, C.E. (1993). Bibliografia e Indici, D. Isella, G. Lucchini e L. Orlando (a cura di), Collana I Libri della Spiga, Garzanti, Milano.

Picchi, E. (1997). DBT 3 - Data Base Testuale: Guida all’uso, versione 3.1. Lexis Ricerche s.r.l. su licenza del C.N.R., Roma.