

Preface

This book corresponds to a (provisional – see the “Conclusion,” Chapter 5) assessment of more of 30 years of work about the possibility of dealing in an “intelligent” and “automated” way with “*narratives*” – in short, with the description of the modalities according to which *some “characters” (not necessarily human) “behave.”* The results obtained up to now in this context, i.e. the Narrative Knowledge Representation Language (NKRL), will be described here in the best possible complete way. Note, however, that, for clarity’s sake, (i) we will avoid as much as possible introducing in the text any cumbersome axiomatic details and, on the other hand, (ii) we have not conceived this book as a practical “manual” of NKRL. We will, nevertheless, supply regularly, in the following chapters, precise indications on where to find the missing “formal” and “operational” information.

All the Web addresses mentioned in the book, both in the text and in the References, were checked for accuracy in November 2007.

The origin of the work described here can be retraced back to 1974, when, after 10 years of professional activity distributed among industry (automation of industrial processes) and academic research at the Centre of Cybernetics and Linguistics Activities of the University of Milan directed by Silvio Ceccato, I was hired as a researcher by the French National Centre for Scientific Research (CNRS) and assigned to a small team of historians and linguists, specialized in the French Middle-Age period and fond of computer science techniques. Given my interests and skills in both computer science and artificial intelligence techniques, two colleagues, Carla Bozzolo and Monique Ornato, suggested I profit from the extensive “biographical” (in the largest meaning of this word) material they had amassed about the main characters of that period to try to build up a computerized system for (i) representing this material in computer memory in the most complete and faithful way; (ii) making use of the stored representations to find all the possible “intelligent” connections among these characters. Thanks to the financial support of the French DGRST (*Direction Générale de la Recherche Scientifique et Technique*, General Direction of the Scientific and Technical Research) the project RESEDA was born (RESEDA: *REseau SEmantique DocumentAire*, Semantic Network for Information Retrieval).

RESEDA's "metalanguage" was far from having the complexity (and the comprehensiveness) of the present NKRL, but many of its fundamental features were already present. We can mention, first, the use of a complex form of knowledge representation going well beyond the usual, simple "object–property–value" models and centered on the use of semantic predicates and roles – the latter surely reminiscent of Ceccato's "correlators." Another important aspect in common with NKRL was the (at least partly) *use of primitives*, both for reducing the risks of combinatorial explosion and for avoiding the ambiguities linked with the intensive use of natural language (NL)-like expressions. RESEDA was, from the beginning, favorably received by the artificial intelligence community [Ornato and Zarri, 1976]; see also the three communications accepted at IJCAI (the International Joint Conference on Artificial Intelligence) in 1977, 1979 and 1981 [King *et al.*, 1977; Zarri, 1979, 1981]. Thanks to additional grants from INRIA (*Institut National de Recherche en Informatique et Automatique*, National Research Institute for Computer Science and Automation) and other French public and private bodies, a complete, running prototype of the RESEDA system was built up, including non-toy knowledge bases of rules and biographical data, a complete backward-chaining inference engine adapted to the particular characteristics of the complex RESEDA's knowledge representation, tools for mass memory storage and update, etc. Given the unavailability at CNRS during the 1970s of advanced symbolic computer tools able to deal with important amounts of data, the system was implemented in VSAPL. On March 2nd, 1984, an official presentation of the complete prototype to the CNRS Scientific Direction put an official end to the project. For a retrospective description of the main technical characteristics of the RESEDA project, see, for example, Zarri [1995] – see also the tribute paid to the originality of this project by J.-P. Genet in his Preface to an "erudite" book about French Middle Age history published recently by Monique Ornato [Ornato, 2001].

The years between 1984 and the early 1990s have mainly been years of theoretical reflection that have seen the progressive transformation of RESEDA's metalanguage into NKRL; a trace of this progressive shift can be found in Zarri [1992a]. We can mention here, for example, the passage from five to seven primitive semantic predicates, with the corresponding increase of the number of exploitable "narrative" structures (what are now known as NKRL's "templates," i.e. the formal description of general classes of events), the final formalization of the specific "sub-language" (AECS) for building up complex arguments of the predicates, the generalization of the "binding structures," etc. Even more important, this restructuration has also implied the addition to the language of a *full, standard ontology* (a hierarchical structure coded according to the classical "binary" approach) in order to represent adequately the concepts and their instances – RESEDA's metalanguage was only endowed with a partially structured "lexicon." With the reorganization of the "catalogue" of templates into a *second ontology* – the "ontology of events," where the nodes (the templates) are represented by *n*-ary structures – NKRL then assumed its

definitive, characteristic configuration based on the contemporaneous presence of two hierarchical structures, i.e. binary for concepts and *n*-ary for templates.

Implementing concretely all the features of the new language in order to produce a complete and running NKRL environment has been a long-term effort, carried out for the most part in the framework of five R&D projects partially financed by the European Commission between 1990 and 2005. Of course, in these 15 years, the “theoretical” developments of the language have been continued in parallel with the implementation work, dealing in particular with the setting up of the particularly sophisticated inference procedures described in Chapter 4.

The first two projects (1991–1995) – NOMOS, Esprit Project 5330, in the legal domain, and COBALT, LRE (Linguistic Research and Engineering) Project 61011, in the corporate domain – have been mainly developed within a generic Common Lisp framework, making use, in case, of specialized environments like CRL (Carnegie Representation Language). A turning point from the viewpoint of the NKRL implementation has been represented by a new Esprit project, CONCERTO (Esprit 29159), which dealt with the advanced management of news stories (1998–2000). The CONCERTO consortium was, in fact, convinced of the necessity to have NKRL implemented in Java to profit fully from its advanced characteristics. A re-implementation in Java (and Java2 successively) of the NKRL modules was then started: at the end of the project (December 2000), a basic Java environment for NKRL was in place, including tools for the setting up of the knowledge bases and a first version of the NKRL *InferenceEngine* modules; see Chapter 4. This work was continued in the subsequent EUFORBIA project (2001–2003), in the “filtering of inappropriate Internet information” domain – EUFORBIA was the 26505 IAP (Internet Action Plan) Project – and finished, at least partially (see the “Conclusion” in Chapter 5 and Appendix A), thanks to the PARMENIDES project (2003–2005) in a “temporal data mining and terrorism news stories” context, Esprit project 39023. Work accomplished within this last project is particularly important for at least two reasons:

- The set up of a complete and efficient implementation of the possibility of executing the NKRL inference procedures in an *integrated way*; see Chapter 4.
- The complete restructuring of the NKRL environment into *two pre-commercial versions*, i.e. a file-oriented and an ORACLE-supported one.

Some “commercial” applications of the NKRL technology have also been implemented during recent years, e.g. in the “beauty care” and “tutoring” domains.

The next, most important moves, see again the “Conclusion” in Chapter 5 for more details, will now concern mainly:

- the setting up of a *complete solution* for accelerating the semi-automatic construction of NKRL knowledge bases through the syntactic/semantic analysis of NL texts – partial results have already been obtained in this context;

- the transformation of the actual prototype(s) into full commercial products – the *first steps* in this direction have already been taken.

Even if I am the main driving force behind both the RESEDA and NKRL projects and the *only one responsible for all the conceptual, technical and implementation choices* accomplished in this context, it is evident that several people have contributed, both from a theoretical and a practical point of view, to the fulfillment of their aims. Besides Carla Bozzolo and Monique Ornato already mentioned – Monique has been especially instrumental in the setting up of the first versions of RESEDA's metalanguage – I would like to mention here the following people: Pedro Abreu, Christophe Assemet, Saliha Azzam, Luc Bernard, Pieter De Vries, Marthe Faribault, Luca Gilardoni, Marie Hayet, Sébastien Hourcaillou, James Henry Ingrouille, Sarah Jacqmin, Margaret King, Dehbia Laradi, Georges Lee, Ruddy Lelouche, Béatrice Marin, Jeremy Martin, John McNaught, Vincent Meissonnier, Bernard Nallet, Fátima Pires, Patricia Sandjong, Lionel Stouder, Peizheng Wu, Lucia Zarri-Baldi, Anne Zwiebel. May the people I have forgotten to quote here forgive me for this omission.

Gian Piero Zarri

Representation and Management of Narrative
Information

Theoretical Principles and Implementation

Zarri, G.P.

2009, X, 302 p. 55 illus., Hardcover

ISBN: 978-1-84800-077-3