

Title	A terminology-based re-definition of grey literature
Author(s)	Claudia Marzi, Gabriella Pardelli, and Manuela Sassi, CNR, Consiglio Nazionale delle Ricerche, Institute of Computational Linguistics, ILC (Italy)
Session One Paper 2	

The Luxembourg Convention on Grey Literature held in 1997 offered the following definition of Grey Literature (expanded in New York, 2004): "Information produced and distributed on all levels of government, academics, business and industry in electronic and print formats not controlled by commercial publishing, i.e. where publishing is not the primary activity of the producing body".

Is this definition still valuable? Is it so far completely satisfactory? Or does it rather need important modifications? We suggest that an interesting re-definition of GL can be based upon careful examination of the longitudinal trend of 10 years of terminological creativity in the proceedings of the GL international Conference. Our empirical basis is the Corpus of GreyText Inhouse Archive, available on <http://www.greynet.org/opensiglerepository.html> consisting of titles, themes, keywords and full abstracts, for a total amount of more than sixty thousand word tokens.

In the full version of our paper, we intend to focus on a set of automatically-acquired terms (both single-word and multi-word terms) obtained by subjecting our reference Corpus to a number of pre-processing steps of automated text analysis, such as concordances, frequency lists and lexical association scores (e.g. Mutual Information on word pairs).

To anticipate some of our results, the following three terms, that appear to be shared by various disciplinary sub-fields, mark, in our view, important stages in the evolution of our current understanding of GL: digital, access and web.

The attribute digital, an increasingly popular synonym of the now obsolete electronic, emphasises the growing importance of computer-based encoding as the standard medium of GL. The noun access (defining the process of accessing text documents) is seen in the company of adjectives like easy, full, grey and open to shape up important conceptual innovations in the way GL material is distributed: e.g. open access focuses on the free accessibility of digital contents. Coupled with information, document and repository, access points to a conception of world-wide available, structured cultural contents. Finally, reference to the web lays emphasis on the huge importance of the World Wide Web as the standard means of disseminating GL.

All these aspects are not fully taken into account in the standard definition of GL reported above. Our inquiry is intended to pave the way to a bottom-up re-definition of GL, stemming from the terminological creativity and lexical innovation monitored over ten years of technical work in the field.

Bionote(s)

Claudia Marzi graduated in Foreign Languages and Literature at Pisa University in 1998, with the dissertation "The power of words: language creativity in Edgar Allan Poe's narrative". Since 2001 she has been working at the CNR Institute for Computational Linguistics in Pisa. She contributed to many national (funded by Italian Ministry of Research) and international (funded by European Commission) projects. Recently she was principal organizer of the European Science Foundation Exploratory Workshop "Words in action: Interdisciplinary approaches to understanding word processing and storage" (Pisa, Oct. 2009)

Gabriella Pardelli graduated in Letters at the University of Pisa in 1980 and has been working at the Institute for Computational Linguistics of the National Research Council in Pisa since 1984. She has been active in the creation of bibliographical databases for Natural Language Processing (NLP) and Digital Library in the Humanities. Other interests regard terminology and History of Human Language Technology. She is responsible for the Library of the Institute of Computational Linguistics and the collection called "Antonio Zampolli Fund". She has participated in many research projects and has worked on the creation of bibliographical resources in the field of language technologies. She has presented many works at different national and international conferences and congresses.

Manuela Sassi graduated in Foreign Languages and Literature at Pisa University, 110/110 cum laude. Since 1974 she has been working in Pisa at the Institute for Computational Linguistics of the National Research Council. Her interests and experiences range from linguistic to textual data processing and in providing linguistic resources on-line. She has been responsible for many national projects and has participated in numerous international projects.