NICOLETTA CALZOLARI
Istituto di Linguistica Computazionale
Pisa, Italy
glottolo@ilc.cnr.it

Language Resources and Content Interoperability. Technical, strategic and political issues for a new generation of Language Resources

Abstract

In the last decade so-called language resources (LRs) (both lexicons and corpora, but also grammars, ontologies, terminologies, and basic and robust software components) have been unanimously recognised as a necessary preliminary platform for developing an adequate Human Language Technology (HLT). Large scale LRs are therefore considered as the infrastructure underlying language technology.

If we broaden our perspective into the future, it is clear that the need of ever growing LRs makes it necessary to propose and promote a change in the paradigm of their creation, management and maintenance. A radical shift in the lexical paradigm - whereby many participants add linguistic content descriptions in an open distributed and collaborative language framework - is required and proposed to make the Web usable. Existing experience in LR development proves that such a challenge can be tackled only by pursuing a truly interdisciplinary approach, and by establishing a highly advanced environment for the representation and acquisition of lexical information, open to the reuse and interchange of lexical data. As a prerequisite to enable this open language infrastructure two key issues are: (i) standards, which are critical to achieve interoperability and integration; (ii) content, which must be dealt with in a multilingual environment and in a dynamic approach.

Another critical issue is the design of general strategies and an overall coordination for the field of LRs as a whole, critical to satisfy some of the requirements of the multilingual information society. My objective is to show that it is important that there is an underlying global strategy behind the set of initiatives which are/can be launched in EU and world-wide, and that a global vision is necessary to achieve more coherent and useful results.

Some introductory issues

The recent past

If we look at the past, in the last decade so-called language resources (LRs) (both lexicons and corpora, but also grammars, ontologies, terminologies, and basic and robust software components) have been unanimously recognised as a necessary preliminary platform for developing an adequate Human Language Technology (HLT) (Calzolari, Zampolli, 1999). Large scale LRs are therefore nowadays considered as the “infrastructure” underlying language technology. In recent years many activities, at European level and world-wide, have contributed to substantially advance knowledge and capability of how to represent, create, maintain, acquire, access, tune, etc. large lexical and textual repositories. These repositories are rich in linguistic knowledge (and often in world knowledge), and many are based on best practices and standards that have been consensually agreed on or have been submitted to the international community as de facto standards. Core - or even large - lexical repositories have been and are being built for many European (and non EU) languages. Many came into existence in European projects, and continued in National Projects, thus creating the necessary platform for a future European language infrastructure.

The role of the EU Spoken and Written LRs projects

The EU Spoken and Written LRs projects of the last decade have achieved the emergence of a broad conscience in EU of the aspects of consensual agreement vs. those of more difficult theoretical or technological solution with respect to the state-of-the-art (Calzolari, 1998). These projects were helpful in creating a more homogeneous community in EU by compelling researchers from different countries and from public and private organisations to work together. They were also effective in spotting and bringing to light a number of commonly felt needs to which some solution had to be found. It was seen as a waste of money, effort and time the fact that any new project started to redo again and again (the same type of) fragments of LRs, without reusing what already available, while LRs produced by the projects - even though small - were usually forgotten and left unused. From here the notion of “reusability” arose (Zampolli, 1987; Calzolari, 1991). The pooling together of these types of requirements made it compulsory, and almost inevitable, the emergence of the concept of LRs as infrastructural types of resource Zampolli (1998).

The strategic infrastructural role of LRs

If LRs - Written, Spoken, and recently Multimodal - are a central and strategic component of the so-called “linguistic infrastructure” (the other key element being Evaluation), necessary for the development of any HLT application and product, the availability of adequate LRs for as many languages as possible becomes a pre-requisite for the development of a truly multilingual Information Society. They play a critical role, as a horizontal technology, in different areas of the EU 6th Framework Programme, and have been recognized as a priority within a number of national projects around Europe. Moreover, ENABLER (European National Activities for Basic Language Resources) (http://www.enabler-network.org/) – an EC funded IST project, designed and started by Antonio Zampolli, with a clear strategic vision for the field of LRs – has recognised the importance to promote actions aiming at integrating these different resource types, until now developed independently, and – as a consequence – to promote the cooperation between the communities of Speech, Text and Multimodality (Zampolli et al., 2000).

The availability of LRs is also a “sensitive” issue, touching directly the sphere of linguistic and cultural identity, but

also with economical, societal and political implications. This is going to be even more true in the new Europe with 25 languages on a par.

The ENABLER Thematic Network of HLT National Projects in European countries was the first broad European initiative with the mission of explicitly considering together the technical, organisational, strategic and political issues of LRs. In ENABLER these various aspects are put together in a coherent framework (Baroni, Calzolari, Lenci, 2003; Calzolari et al., 2004), to set up a medium- and long-term set of priorities (both technical and strategic) and to promote these at the national and international levels.

Looking at the future

If we broaden our perspective into the future, it is clear that the need of ever growing LRs – testified also by the current US funding strategies – makes it necessary to propose and promote a change in the paradigm of their creation, management, maintenance, and sharing. A radical change of perspective is now needed in order to facilitate the integration of the linguistic information resulting from all the LR initiatives of the last years. We need i) to bridge the differences between various perspectives on language structure and linguistic content, ii) to put an infrastructure into place for content description and content interoperability at European level and beyond, and by this iii) to make lexical and language resources usable within the emerging Semantic Web scenario (Calzolari, 2002).

This objective can only be achieved when working in the direction of an integrated Open and Distributed Language Infrastructure, based on open content interoperability standards, where not only the linguistic experts can participate, but which includes designers, developers and users of content encoding practices, and also many members of the society. A radical shift in the lexical paradigm - whereby many participants add linguistic content descriptions in an open distributed and collaborative language framework - is therefore required and proposed to make the Web usable. Existing experience in LR development proves that such a challenge can be tackled only by pursuing – on the organizational side - a truly interdisciplinary and cooperative approach, and by establishing – on the technical side - a highly advanced environment for the representation and acquisition of linguistic information, open to the reuse and interchange of linguistic data.

The new proposed paradigm of an Open Language Infrastructure requires new approaches at various levels. The effort of making available millions of ‘words’ for dozens of languages is something that no individual or small group is able to afford. On the other side, it is already proved by a number of international projects that lexicon building and maintenance can be achieved in a cooperative way. We claim that the field of computational lexicons and of LRs in general is mature enough to broaden and open the concept of cooperative effort to a much larger set of communities, so that lexicon creation, updating and maintenance will be no longer only a lexicographers’ task, but will involve broad groups of experts, and – why not? - also ‘non-experts’ and the general public.

LRs and the Semantic Web vision

Language - and lexicons - are the gateway to knowledge: only through them can we tackle the twofold challenge of digital content availability and multilinguality. Aiming at making word content machine-understandable, computational lexicons intend to provide an explicit representation of word meaning, so that it can be directly accessed and used by computational agents, such as large-coverage parsers, modules for intelligent Information Retrieval or Information Extraction, etc. In all these cases, semantic information is necessary to enhance the performance of Natural Language Processing tools, and to achieve a real understanding of text content. Moreover, in multilingual computational lexicons we find the linguistic (morpho-syntactic/semantic) information necessary to establish links among words of different languages, information of great importance for systems performing multilingual text processing, such as Machine Translation, Cross-lingual Information Retrieval, etc.

Semantic content processing lies at the heart of the Semantic Web vision. Semantic Web developers will therefore need very large repositories of words and terms - and knowledge about their relations. The cost of adding this structured and machine-understandable lexical information can be one of the factors that delays Semantic Web full deployment. A natural convergence thus exists between some of the core activities in the field of HLT and the Semantic Web long-term goals.

Critical issues for the new paradigm

The approach to realise such a distributed language infrastructure requires the coverage not only of a range of aspects pertaining to linguistic modelling, but also – and maybe most critically – of a number of organisational aspects. Moreover, such a language infrastructure is inherently market driven, since the most widely used LR portions will be the best developed and supported.

To facilitate the integration of the LRs and tools resulting from all the various LR initiatives of the last decade and, at the same time, to make word-content machine understandable, as it is the aim of the Semantic Web vision, three critical issues must be addressed:

standards, critical to achieve the interoperability needed for effective integration;

content, as crucial information to be represented is semantic information which must be dealt with in a dynamic approach;

multilinguality, a critical issue for the immediate future.

Answers to these issues are found within two frameworks: (i) in the context of the ISLE (Calzolari et al., 2002) enterprise (http://www.ilc.cnr.it/EAGLES96/isle/ISLE_Home_Page.htm) which, with the definition of MILE (Multilingual ISLE Lexical Entry), represents an interface between advanced research in multilingual lexical semantics and the practical task of developing resources for HLT; (ii) in lexicon models that account for the complex, multidimensional and multifaceted nature of meaning in lexicon and ontology design, and can accommodate

dynamically acquired information in a multilingual environment.

A few keywords/desiderata for an Open Lexical Infrastructure are: open framework, multilingual, multimodal, multimedial, dynamic, integrative, knowledge intensive, consensual, interactive. We will briefly expand here on some of these.

Multilinguality

Multilinguality is seen by ENABLER as a critical issue in the information society, and in particular in Europe, and most of its recommendations are linked, in one way or another, to it. Multilinguality is also a strong integrating factor, being horizontal with respect to different application areas and to different LR types (Spoken and Written), and implies not only harmonised technical decisions, but also heavy organisational aspects, which can only be taken care at supranational level. A way to reach the optimisation of the process of production and sharing of (multilingual) LRs can be found in a common and standardized framework which ensures the encoding of linguistic information in such a way to grant its reusability in different applications and tasks. Many European initiatives have built large-scale monolingual LRs and lexicons for many languages on the basis of common specifications and shared theoretical models, as a prerequisite for their multilingual linking.

Open access to LRs

A Declaration on Open Access to LRs was issued at the ENABLER/ELSNET Workshop “International Roadmap for Language Resources” (Paris, August 2003, for which see http://www.enabler-network.org/final-workshop.htm) and endorsed by all the participants. With respect to this issue we have set up a virtual network, comprising the major LRs and HLT groups in Europe and world-wide, ready to launch a large initiative to open LRs on the Web. The outcome of such initiative could be the design of a completely “new generation” of LRs.

Standards for content interoperability: towards lexicon and LR integration

The ENABLER Network promoted the compatibility and interoperability of LRs mainly through cooperative work with: i) ISLE/EAGLES, for harmonisation of linguistic specifications, in particular for corpora and multilingual lexicons (based on MILE); ii) ISO TC37 SC4 WG4 (http://www.tc37sc4.org/WG4/wg4.htm), to make European standards truly international ISO Standards; iii) ELRA Validation Committee (http://www.elra.info/article.php3?id_article=44) , for the incorporation of agreed standards in protocols for the validation of LRs, both Spoken and Written; iv) INTERA (http://www.elda.org/rubrique22.html), for the harmonisation of metadata descriptions; v) Semantic Web communities, to promote synergy between the groups of knowledge management/ontology and of HLT/LR technology.

The model of open data categories will foster LR integration and interoperability, through links to common standards. With the ISLE approach to lexical standards, and its definition of the MILE (Multilingual ISLE Lexical Entry) (Calzolari et al., 2003), new lexical objects can be progressively created and linked to the core set.

The reuse of existing lexicons - and the sharing of lexical resources in general - is achieved through the design of MILE, a lexical schema which (i) takes into consideration the basic notions employed in major available lexicons, (ii) is flexible enough to allow mapping from various lexical models into it, and (iii) allows the creation of user defined lexical objects if needed. The multidimensional perspective is one of the peculiar features of the ISLE activities, and contributes to its added value with respect to other current standardisation initiatives. This way, ISLE intends, on the one hand, to answer to the need of fostering the reuse and interchange of existing lexical resources and, on the other hand, to enhance the technological transfer from advanced research to applications. It also prepares the ground for a “new generation” of “knowledge resources”. The design of an abstract model of lexicon architecture will ensure a flexible model while working with a core set of lexical data categories. It will guarantee freedom for the user to add or change objects if that is deemed necessary, will provide an evaluation protocol for the core standard lexical data categories, and will require verification methods for the integration of new objects.

We foresee an increasing number of well-defined linguistic data categories and lexical objects stored in open and standardised repositories which will be used by different types of users to define their own structures within an open lexical framework. This is currently dealt with in the ISO TC37 SC4 WG4 for Lexical Resources. It is this re-usage of shared linguistic objects which will link new contents to the already existing lexical objects, while enabling shareability of distributed lexicon portions.

From “static” towards “dynamic” LRs

The promotion of a change of perspective on lexicons as static resources towards dynamic entities, whose content is co-determined by automatically acquired linguistic information from text corpora and from the web, is deemed essential. We stress the need of using (semi-)automatic or machine-aided methods wherever possible in LR work.

In automating the information extraction process, lexical resources are one fundamental piece, and acquisition may require cycles of adjustments in the lexical modelling. The acquisition tools must be able to increase the repository with new words/terms, possibly their definitions, domain, etc., from digital material, to learn concepts from text – including automatic multilingual thesaurus building, and to tailor resources to specific needs. This implies, for the machine learning community, developing new and stronger algorithmic methodologies to model textual statistics, and integrating them with traditional NLP tools. Agents will look for examples, identify uses in monolingual/multilingual web texts for glossary creation. This will ensure also virtual links between lexicons and examples (corpus/web samples, image samples, clips and videos, etc.), pushing towards new and innovative types of lexicons, i.e. a sort of ‘example-based living lexicons’, linguistic objects that participate of properties of both lexicons and corpora.

Computational lexicons will be more and more conceived as dynamic systems, whose development needs to be complemented with the automatic acquisition of semantic information from texts. Gaining insights into the deep

interrelation between representation and acquisition issues is likely to have significant repercussions on the way linguistic resources will be designed, developed and used for applications in the years to come. As the two aspects of knowledge representation and acquisition are profoundly interrelated, progress on both fronts can only be achieved through a full appreciation of this deep interdependency.

Semantic Web initiatives are also focussing on the building of ontological representations from texts, and in this respect show a large amount of conceptual overlap with the notion of a dynamic lexicon.

The new paradigm of an open and distributed language infrastructure

The proposed open and distributed language infrastructure must be able to simultaneously tackle the following aspects:

the design of advanced architectures for the representation of lexical content;

the development of new methods and techniques for the automatic acquisition of semantic knowledge from texts and for the customization and update of lexical resources;

the standardization of various aspects of the lexicon, up to content interoperability standards.

Some of the prerequisites of this new lexical framework are:

open framework, where everyone must be able to access, put new information, get parts of the lexicons;

multilingual, multimodal, multimedial, and dynamic i.e. comprising tools for acquiring information from different media (e.g. on the web);

integrative, allowing the integration into different environments and enabling an interaction between different types of LRs (e.g. corpus and lexicon);

knowledge intensive, allowing representation of rich semantic information, and bootstrapping new semantic information starting from the available one.

Coming from the experience gathered in developing advanced lexicon models such as the SIMPLE one (Ruimy et al., 2003), and along the lines pursued by the ISLE standardisation process, a new generation of lexical resources can be envisaged. These will crucially provide the semantic information necessary to allow for effective content processing. On the other hand, they will in turn benefit from the Semantic Web itself. Thus, it is possible to state the existence of a bidirectional relation between the Semantic Web enterprise and computational lexicon design and construction. In fact, the Semantic Web is going to crucially determine the shape of the LRs of the future. Semantic Web emerging standards, such as ontologies, RDF, etc., allow for a new approach to LR development and maintenance, which is consistent with the vision of an open space of sharable knowledge available on the Web for processing.

We claim that it is feasible and timely – and indeed critically needed - to set up a very large world-wide initiative, within which a large group of lexicon experts/providers is a core component. This initiative must build prominently on existing or future (possibly national) lexical initiatives, and integrate along this dimension the human R&D resources coming from both the communities of HLT and Semantic Web. This may become the new 'vision' for LRs in the next years.

A cooperative model

This vision, enabled by MILE, will pave the way to the realisation of a common platform for interoperability between different fields of linguistic activity - such as lexicology, lexicography, terminology, corpus linguistics - and Semantic Web development. The platform will provide a flexible common environment not only for linguists, terminologists and ontologists, but also for content providers and content management software vendors. The lexicons may be distributed, i.e. different building blocks may reside at different locations on the web and are linked by URLs. This is strictly related to the Semantic Web standards (with RDF metadata to describe lexicon data categories), and will enable users to share lexicons and collaborate on parts of them.

Overall, the lexicons will perform the bridging function between textual documents and conceptual categorisation. The descriptive framework and the tools will be tailored to the needs of the creation of both general vocabularies and terminological repositories for different worlds and domains. The common conceptual model within the envisaged architecture will ensure content interoperability between texts, multimodal documents, lexicons and ontologies.

Strategic issues and international research infrastructures for S&W LRs

Such an infrastructure is part of a more global strategic vision for the field of LRs, within which international cooperation will be certainly the most important factor for LRs – and consequently for HLT - in the next years. ENABLER has been particularly active in this respect, working towards designing an overall strategy for LRs (Mapelli, Choukri, 2003b). The goals were: to provide recommendations for strategic initiatives to be promoted for LR production and management, to address the main priorities for LRs, to start designing, in cooperation with ELSNET, a roadmap for LRs, and to define a strategy for LRs in the next years. Two main lines have been highlighted:

infrastructural initiatives – ENABLER has promoted the creation of a new international infrastructure for LRs;

coordination initiatives – these concern both the national and the transnational and transcontinental dimensions.

We briefly highlight here some of these initiatives (for more details see http://www.enabler-network.org/).

A common Roadmap for S&W LRs and HLT

The workshop "International Roadmap for Language Resources" (http://www.enabler-network.org/final-workshop.htm), jointly organised by ENABLER and ELSNET (http://www.elsnet.org/) in Paris in August 2003, has actually laid the basis to build a roadmap for LRs. Another

Roadmap Building meeting, as a common enterprise of the spoken and written communities, was held in conjunction with LREC 2004 in Lisbon (http://www.lrec-conf.org/lrec2004/). A first list of priorities which act as critical issues for the future of both spoken and written LRs was drawn:
provide basic LRs for a larger set of languages;
increase multilingual LRs;
reduce development time of LRs;
enhance LR content interoperability;
foster synergies between spoken and written areas and with neighbouring areas (e.g. terminology, Semantic Web);
develop new methodologies and tools for LR management, quick domain and application adaptation, data-driven tuning, etc.

These initiatives are helpful also to build a summary of the technical, operational and organisational problems to be tackled, and to provide suggestions for an overall organisation framework for international cooperation.

BLARK (Basic LAnguage Resource Kit)

The BLARK (Basic LAnguage Resource Kit) concept, first launched through ELSNET (Krauwer, 1998) and Nederlandse Taalunie (Binnenpoorte et al., 2002), was promoted and expanded by ENABLER through i) the definition of a minimal set of LRs (in terms of spoken and written corpora, lexicons, basic tools to manipulate them, skills required, etc.) to be made available for as many languages as possible, to be able to do any pre-competitive research for that language; and ii) the spotting of the actual gaps to be filled in order to meet the needs of the HLT field. For this a BLARK matrix is being implemented at ELRA (http://www.elra.info/), to highlight the gaps with regard to the LRs needed for specific applications and for as many languages as possible. This matrix will be made accessible and modifiable directly from the ELRA Web site, to enable customers or providers of LRs to fill it in with complementary information, and help ELRA to identify available LRs and to promote the production of new specific LRs. Such an initiative, combined with other ongoing actions, will contribute to map and, in the end, to fill a fair number of the gaps, and thus to improve the working material of the HLT community. Among these initiatives, we should not omit the maintenance and updating work on LRs, further to the production work, since the cost of LRs is high enough to take into consideration their reusability on a long-term basis.

Moreover, BLARK must be considered as an evolving notion. A further level is defined as Extended LAnguage Resource Kit (ELARK), which will be extensively promoted for its larger adoption (Mapelli, Choukri, 2003a).

International cooperation

In order to fill the gaps in terms of LRs, cooperation on all combined - organisational, funding, technical and commercial - issues appears to be necessary. To strengthen such a cooperation, there is no doubt that an effort in coordinating this cooperation is required. The major strategic outcomes of ENABLER with respect to international cooperation and to the design of an overall coordination and strategy in the field of LRs are: i) the founding of the International coordination Committee for Written Language Resources and Evaluation (ICCWLRE) that will work together with COCOSDA for Spoken LRs, and ii) the preparation of a proposal – LangNet - to be submitted within the ERA-Net scheme of the 6th Framework Programme for the coordination of national activities in the area of HLT.

WRITE

A coordinated operation was already launched in the framework of Speech LRs with the creation of COCOSDA (International Committee for the coordination and Standardisation of Speech Databases and Assessment Techniques).

A new committee, originally conceived by Antonio Zampolli, has been established in the field of Written LRs, the Written Resources Infrastructure, Technology and Evaluation (WRITE) Committee (http://www.ilc.cnr.it/write/), which is the natural evolution of ICCWLRE. It provides the optimal environment to continue (part of) the ENABLER mission, while, at the same time, enlarging its scope beyond the European boundaries. Tasks for this Committee include: information dissemination on LRs and standards; promotion, coordination, and enabling activities; copyright and IPR; training and methodology for LR creation and validation; roadmaps for LRs; political and strategic initiatives.

LangNet

Last but not least, an initiative – LangNet – is being proposed in the framework of the ERA-Net scheme of the 6th FP of the European Commission (EC). The objective of the LangNet proposal is to establish some sort of permanent coordination among national initiatives in HLT all over Europe, to capitalise on parallel national initiatives on the long run. The LangNet initiative thereby candidates itself to provide the most natural environment to continue the efforts and the momentum gained by the ENABLER Network. Language Technologies seem to be especially well fitted for the ERA-Net scheme, based on the assumption that each country wishes to conduct research activities allowing the development of systems and applications for its language(s). It therefore seems natural that the individual countries basically take into account all the “(spoken and written) language-dependent” aspects and that the EC rather takes into account all the generic, “language-independent” aspects, in agreement with the principle of subsidiarity. The linguistic infrastructure supported by ENABLER and proposed by LangNet intends to contribute to the structuring and integration of the European Research Area, addressing problems such as the fragmentation of its research base and the weakness in converting R&D results into useful economic or society benefits.

As pointed out very well in the final Euromap Report (Joscelyne, 2003), coordination initiatives must be put in place so as to avoid a two-speed situation, between languages which are interesting commercially, and today also politically (for defence, intelligence, etc.), and those which are not (unfortunately the vast majority). This implies for Europe that coordination should be established between the European Commission and the member states and strategies should be drawn in order to ensure a proper balance of language coverage in Europe.

Conclusion

In order to set up the required world-wide language infrastructure on the web, however, an essential aspect for ensuring an integrated basis is to enhance the interchange and cooperation among many communities that act now separately, such as Language Resources developers, Terminology, Semantic Web and Ontology experts, content providers. This is the challenge for the next years, for a usable and useful lexical/linguistic scenario in the global network.

At the end everything is tied together, which makes our overall task so interesting - and difficult. What we must have is the ability to combine the overall view with its decomposition into manageable pieces. No one perspective - the global and the sectorial - is really fruitful if taken in isolation. To this aim, we claim it is necessary to pool together and to build on many different, but related, initiatives. A strategic and visionary policy for cooperation between different groups has to be debated, designed and adopted for the next few years, if we hope to be successful, but - inside this - a realistic and stepwise approach to solving well-defined and limited aspects must be adopted. To this end, the contribution of the main actors in the field is of extreme importance. Some of the events of the last years are hopefully moving in this direction.

References

1. Baroni, P., Calzolari, N., Lenci, A. 2003. Extended Configuration of the Network and Final Report. ENABLER Deliverable D1.2, Pisa.
2. Binnenpoorte, D., De Vriend, F., Sturm, J., Daelemans, W., Strik, H., Cucchiarini, C. 2003. “A Field Survey for Establishing Priorities in the Development of HLT Resources for Dutch”. In LREC 2002 Proceedings. Las Palmas, pp. 1862-1866.
3. Calzolari, N. 1991. “Lexical databases and textual corpora: perspectives of integration for a Lexical Knowledge Base”, U. Zernik (ed.), Lexical Acquisition: Exploiting on-line Resources to build a Lexicon. Lawrence Erlbaum Associates, Hillsdale, NJ, pp. 191-208.
4. Calzolari, N. 1998. “An overview of Written Language Resources in Europe: a few reflections, facts, and a vision. In A. Rubio, N. Gallardo, R. Castro, A. Tejada (eds.), Proceedings of the First International Conference on Language Resources and Evaluation. Granada, 217-224.
5. Calzolari, N. 2002. “Computational Lexicons: Towards a New paradigm of an Open Lexical Infrastructure?”. In G. Willée, B. Schröder, H.C. Schmitz (eds.), Computerlinguistik. Was geht, was kommt?. Computational Linguistics. Achievements and Perspectives. Gardez! Verlag, Sankt Augustin, pp. 41-47.
6. Calzolari, N. 2004. “Introduction of the Conference Chair”. In LREC 2004 Proceedings. Lisbon.
7. Calzolari, N., Bertagna, F., Lenci, A., Monachini, M. (eds.). 2003. Standards and Best Practice for Multilingual Computational Lexicons. MILE (the Multilingual ISLE Lexical Entry). ISLE CLWG Deliverable D2.2&D2.3. Pisa, pp. 194. http://lingue.ilc.cnr.it/EAGLES96/isle/.
8. Calzolari, N., Choukri, K., Gavrilidou, M., Maegaard, B., Baroni, P., Fersøe, H., Lenci, A., Mapelli, V., Monachini, M., Piperidis, S. 2004. “ENABLER Thematic Network of National Projects: Technical, Strategic and Political Issues of LRs”. In LREC 2004 Proceedings. Lisbon, pp. 937-940.
9. Calzolari, N., Lenci, A., Bertagna, F., Zampolli, A. 2002. Broadening the Scope of the EAGLES/ISLE Lexical Standardization Initiative. In N. Calzolari, K. Choi, A Lenci, T. Tokunaga (eds.), Proceedings of the 3rd Workshop on Asian Language Resources and International Standardization. Taipei, Taiwan, pp. 9-16.
10. Calzolari, N., Zampolli, A.. 1999. “Harmonised large-scale syntactic/semantic lexicons: a European multilingual infrastructure”. In MT Summit Proceedings. Singapore.
11. Joscelyne, A., Lockwood, R. 2003. Benchmarking HLT Progress in Europe. HOPE, Copenhagen.
12. Krauwer, S. 1998. “ELSNET and ELRA: A Common Past and a Common Future”. In ELRA Newsletter, 3(2).
13. Mapelli, V., Choukri, K. 2003a. Report on a (Minimal) Set of LRs to Be Made Available for as Many Languages as Possible, and Map of the Actual Gaps. ENABLER Deliverable D5.1, Paris.
14. Mapelli, V., Choukri, K. 2003b. Report Contributing to the Design of an Overall coordination and Strategy in the Field of LRs. ENABLER Deliverable D5.2, Paris.
15. Ruimy N., Monachini, M., Gola, E., Calzolari, N., Ulivieri, M. Del Fiorentino, M.C., Ulivieri, M., Rossi, S. 2003. “A Computational Semantic Lexicon of Italian: SIMPLE”. In A. Zampolli, N. Calzolari, L. Cignoni (eds.), Computational Linguistics in Pisa. Linguistica Computazionale. Vol. XVI-XVII. Pisa, IEPI, pp. 821-864.
16. Zampolli, A. 1987. “Perspectives for an Italian Multifunctional Lexical Database”. In Linguistica Computazionale. Vol. IV-V, Pisa, pp. 304-341.
17. Zampolli, A. 1998. “Introduction of the General Chairman”. In A. Rubio, N. Gallardo, R. Castro, A. Tejada (eds.), First International Conference on Language Resources and Evaluation, Granada, ELRA, pp. xv-xxv.
18. Zampolli, A. et al. 2000. ENABLER Technical Annex, Pisa.