

Curriculum Vitae

Giuseppe Attardi

Address:

Dipartimento di Informatica
Università di Pisa
largo B. Pontecorvo 3
I-56127 Pisa, Italy
tel: +39 (050) 221-2744
cel: +39 (348) 700-4820
mail: attardi@di.unipi.it
<http://www.di.unipi.it/~attardi>

Personal Data:

Born in Padova, 9 June 1950
Citizenship: Italian
Marital status: married, one child



Interests

I wish to dedicate myself as a public servant to advance technologies for improving interactions with computers, building services that help improving the life of citizens and fostering economic development.

I work mostly as a software architect, who designs and develops state of the art software systems to solve challenging problems, advancing the state of the art in computer science.

I have a long standing interest in issues of Knowledge Representation and reasoning, for which I developed formalisms (the Omega description logic [13]), tools (ECL) and programming techniques (CLOS for object-oriented programming).

Currently I explore using text documents as a source for extracting knowledge, by means of document analysis techniques which combine statistical and semantic analysis, including Natural Language Processing (language is still the most effective human mean for expressing and communicating knowledge).

Question Answering is a challenging benchmark application for several of these techniques, from Information Retrieval (using the IXE search library), to machine learning (e.g. Maximum Entropy, Multilayer Perceptron and Named Entity Recognizer within the PiQASso QA system), to Natural Language parsing (the DeSR multilingual dependency parser).

Such challenging applications often require advances in software technologies, so for instance I am interested in generative programming tools for automating certain phases of code generation, in particular for producing optimized code for specific cases.

I am developing Machine Reading technologies and I apply them to tasks such as *opinion mining* [36], identifying shared interests among people and medical record analysis.

Education

Specialization Diploma in Computer Science (Specializzazione in Calcolo Automatico), Università di Pisa, Italy, 1973.

MS in Computer Science (Laurea in Scienze dell'Informazione), Università di Pisa, Italy, 1972.

Academic Career

Full Professor of Computer Science, Università di Pisa, Italy	2001 – present
Senior Scientist, Yahoo! Research Barcelona, Barcelona, Spain	2006 – 2007
Senior Scientist, Sony Research Laboratory, Paris, France	1996
Senior Scientist, International Computer Science Institute, Berkeley, USA	1993
Associate Professor of Computer Science, Università di Pisa, Italy	1986 – 2001
Visiting Scientist, Artificial Intelligence Laboratory, MIT, USA	1978 – 1981
Assistant Professor of Computer Science, Università di Pisa, Italy	1977 – 1986
Research Assistant of Computer Science, Università di Pisa, Italy	1973 – 1977

Software Achievements

- Top score at the CLEF 2012 Pilot Task on Question Answering for Machine Reading on Alzheimer's disease *September 2012*
- Top score at the SuperSense tagging task at Evalita 2011 *January 2012*
- Top score at the POS tagging task at Evalita 2011 *January 2012*
- The DeSR parser achieved second best score on Italian in the Evalita 2011 Dependency Parsing task *November 2011*
- The DeSR multilingual dependency parser achieved best score on Hindi in the ICON 2010 tools contest *December 2010*
- The TanI Anaphora tagger achieved top scores in several languages at SemEval 2010 Task-1 on Coreference Resolution in Multiple Languages *June 2010*
- The DeSR multilingual dependency parser achieved best scores in the Evalita dependency parsing Task *November 2009*
- The DeSR multilingual dependency parser achieved best statistical parser in the Evalita dependency parsing Task *September 2007*
- The DeSR multilingual dependency parser achieved second best accuracy in the Domain Adaptation track of the CoNLL-X Shared Task *April 2007*
- IXE was ranked as the fastest in search among a group of Open Source Search Engines in the thesis by Christian Middleton *March 2007*
- My opinion mining system achieved third p@5 score in the TREC Blog Track *November 2006*
- The DeSR multilingual dependency parser achieved better than average accuracy at CoNLL-X Shared Task *March 2006*
- IXE search engine achieved close second best P@10 score, with fastest performance at TREC 2005 TeraByte task *November 2005*
- PiQASso Question Answering System in top 10 TREC QA task *November 2001*
- CMM garbage collector deployed by Sun Microsystems in the implementation of Java *1996*

Achievements

1974	early paper proposing object-oriented programming methodology
1976	design of a RISC style architecture for object-oriented languages
1978	development of denotational semantics for concurrent actor language
1979	participation in the development of first MIT window system (with Richard Stallman)

- 1979 first graphical spread-sheet (on the MIT Lisp machine)
- 1979 first description logic for Knowledge Representation (Omega)
- 1981 development of formal semantics for semantic network formalism
- 1981 first installation of UNIX 4.1 BSD outside the USA (with Bill Joy)
- 1982 deployment of first heterogeneous Ethernet network in Italy (CSELT, Torino)
- 1983 design and deployment of first campus-wide fiber optic network in Italy (European Joint Research Center), including development of Unix driver for token-ring network adapter
- 1984 realization of largest network of workstations outside the USA (CRAI, Cosenza)
- 1984 development of theory of "*viewpoints*" for reasoning with multiple agents
- 1984 leader of a European research project (MADS P440) selected as flagship ESPRIT project
- 1986 development of a reflective, multithreading scripting language (Abstract Common Machine) for programming software agents across the network [18]
- 1986 design of the thread mechanism of the Chorus Operating System, later purchased by Sun Microsystems
- 1993 development of customizable garbage collector for C++ (CMM) [6]
- 1994 CMM is used by Sun Microsystems in the development of Java
- 1996 development of Arianna (with Italia Online), the first large search engine for the Italian Web
- 1997 design of Italian National Broadband Network for Research (GARR-B)
- 1998 deployment of private city-wide fiber-optic network for the University of Pisa (REALTA)
- 1998 paper "Categorization by Context", describing one of the earliest uses of Web link analysis, received Best Paper Award at WebNet'98 Conference [8]
- 2001 design of Italian National Gigabit Network for Research (GARR-G)
- 2002 invited by Microsoft at Redmond to demonstrate IXE search technology on a search engine built for site microsoft.com.

Business Experience

Founder and Scientific Consultant, WebSays, Spain *July 2010*

- WebSays (websays.com) collects data from multiple sources and provides reports on what is being said that is relevant to a customer.

Founder and Scientific Consultant, Kiwi Srl *July 2010*

- Kiwi develops a localized social network for connecting with people nearby and exchanging messages.

Founder and Scientific Consultant, Comprendo Srl *March 2010*

Founder and Director of Technology, Ideare SpA *January 1999*

- Ideare has been a leader in search technology in Italy, building search engines for several major portals: Italia Online (www.arianna.it), SuperEva, MonrifNet, Tiscali, La Repubblica (the largest national newspaper)

- The company was acquired by Tiscali and developed search facilities for Tiscali in several countries: Italy, Switzerland, France, Germany, and England.

Founder and Director of Technology, DELPHI SpA

January 1982 – December 1989

- The company was acquired by Olivetti in 1987.

Public Administration

Consiglio Tecnico Scientifico of Consortium GARR	2012
Comitato del Polo Informatico Fibonacci, Università di Pisa	2012
Evaluation Committee for PRIN 2006.	
Consiglio Tecnico Scientifico of Consortium GARR	1997-2006
Commissione Rete della CRUI	1996-2001
Direttore del Centro di Calcolo, Dipartimento di Informatica	1989-1994

Current Research Projects

- 2012 MaRea, Machine Reading from multiple public sources, funded by Regione Toscana and Tiscali.
- 2010 PARLI: Strumenti automatici di annotazione sintattica e validazione collaborativa. PRIN 2008.
- 2009 Multilanguage semantic Access to Italian cultural contents on the Web. FIRB 2006.

Recent Research Grants

- 2008 Linguistic Document Analysis for Web Mining and Question Answering. Yahoo! Research donation.
- 2006 Fondazione Cassa di Risparmio, Text Analysis for Semantic Web and Question Answering.
- 2006 Mobile Guides. PRIN (Research Project of National Relevance) 2005.
- 2003 "Parallel Question Answering". Progetto Strategico MIUR Società dell'Informazione 2000.
- 2003 ST Microelectronics: development of graphics technologies for embedded processors.
- 2002 Rotor grant from Microsoft Research.
- 2002 [AppSem II](#), Applied Semantics II, EU 5th Framework Program Thematic Network.
- 2001 KSolutions SpA: development of technologies *Best Bets* query matching.
- 2000 "Technologies for Enhanced Content Delivery". Progetto Strategico MIUR Società dell'Informazione, 1999.
- 2000 Web Switching, [Piattaforme Abilitanti per Griglie Computazionali ad Alte Prestazioni](#). FIRB 2002.
- 2000 Ideare SpA (Tiscali group): development of search engine technologies.

Recent Research Activities

My recent research activities have focused on the following themes: generative programming, document analysis techniques for search and Question Answering.

Document Analysis, Search and Question Answering

Indexing and Search

IXE is a high performance C++ class library for indexing and search on large document collections. The library has been designed with goals of efficiency and scalability, in order to handle document collections of petabyte size. IXE has been used in several commercial applications: Web search engines, multimedia search services, enterprise search. In particular IXE has been powering search for Tiscali (www.tiscali.it) and for the largest Italian national newspaper La Repubblica (www.repubblica.it).

IXE provides a wide range of facilities for document handling, from tokenization to regular expressions, to multiple encodings, as well as a pluggable document-reader architecture. IXE allows creating high throughput search services by means of integrated threading and HTTP support. The library has been designed for implementing high performance full-text search services, but is capable of handling more complex structures than pure text documents by means of an object store that provides persistency to C++ objects.

Query processing in IXE is performed through a generalized cursor interface, which provides a single operator adequate for handling any kind of query (boolean, proximity, multiple collection, etc.). IXE exploits data structures and algorithms optimized for each task: for instance the search for multiple keywords uses the algorithm Small Adaptive Set Intersection, which was proved optimal for this problem.

IXE has achieved top query performance at TREC 2005, corresponding to over 20 queries per second on a collection of 20 million documents on a single Pentium PC.

Question Answering

Question Answering involves a combination of techniques of *information retrieval*, *natural language processing*, *machine learning*, *information extraction*, in order to provide precise answers to queries expressed in natural language.

PiQASso is a full question answering system, which ranked among the top ten at TREC 10 [29]. PiQASso performs a syntactic analysis of the question by means of a dependency parser, identifies the answer type and performs a search on a Passage Retrieval engine, extracting relevant paragraphs from the document collection; each candidate answer paragraph is parsed and a series of filters are applied to discard sentences which do not appear to contain an adequate answer. These filters analyze:

- the expected answer type;
- the structure of the sentence, computing a matching distance between semantic correspondences in parts of the question and of the answer;
- popularity of the answer.

The semantic structure matching approach was later adopted by National University of Singapore ranking second best in later TREC QA conferences.

The latest version of PiQASso incorporates Named Entity tags in the document index, so that more precise paragraphs can be selected in the retrieval stage.

Semantic Search

The technology of inverted indexes is quite effective to achieve fast query processing. *DeepSearch* is a semantic search engine which exploits inverted indexes in an original way. A semantically annotated document can be considered as consisting of multiple layers where each layer represents a level of annotation, e.g. POS, lemma, named entity, semantic category, dependency. The inverted index stores

information of tokens occurring at each position in a document. Multiple indexes can represent multiple layers and with a special *Align* query operator can express attributes of word occurring across documents. The efficiency of indexes allows answering such complex query at no extra cost. For the TREC Blog Opinion Mining task I used such an index were a layer contained annotations for opinionated words, allowing me to produce the fastest system among all submissions. *DeepSearch* (<http://semawiki.di.unipi.it/search/demo.html>) is an application of this technique for searching the Wikipedia.

Natural Language Parsing

The linguistic structure of sentences can be described in terms of dependencies between words, instead of phrase structure trees, as in classical linguistics. The *dependency* relation is an asymmetric relation between a *head word* and a *dependent word*. A dependency has an associated type, expressing the grammatical function of the dependency (Subject, Object, Predicate, Determiner, Modifier). In a dependency tree, one word is the head of a sentence, and all other words are either a dependent of that word or else a dependent on some other word which connects to the headword through a sequence of dependencies. Word-word dependency structure is far easier to understand and more convenient to use by further analysis tools.

There are advantages of dependency grammars also for parsing. Because dependency grammars work directly in terms of dependencies between words, disambiguation decisions are made in terms of these word dependencies. Another advantage of dealing with dependencies is that they decompose the sentences into pairs, i.e. smaller and more modular units than trees with several children. Smaller units will occur more frequently and therefore statistical learning algorithms can rely on a larger number of samples and hence with more statistical relevance.

I have built a multilingual statistical dependency parser [34] based on the approach of transition-based parsing. The parser is capable of handling non-projective relations and achieves state of the art accuracy. The algorithm has linear complexity so that it can process over 50,000 tokens per second.

I developed a patent-pending technique for correcting parser errors based on learning tree revision rules [37]. Besides improving the accuracy of dependency parsers, the technique has proved effective in domain adaptation, i.e. adapting a parser, trained on a corpus in one domain, to analyze text from another domain for which only unannotated data is available [38].

Best Bets

A number of applications require selecting targets for specific contents on the basis of criteria chosen by the contents providers themselves: advertising services (e.g. Googles AdWords, AdSense and sponsored links), editorial promotions, shopping advice, recommending systems, matchmaking applications in e-commerce. Best Bets is a class of retrieval systems that generalize Information Filtering and encompass these kinds of applications. Implementing Best Bets involves inverting the direction of search, retrieving queries from documents rather than vice versa, as in ordinary information retrieval. Because of this, normal inverted indexing and search techniques cannot be applied, and special data structures and techniques have been developed.

We have addressed performance issues in the implementation of Best Bets [32], in particular efficient search, incremental updates and dynamic ranking.

Generative Programming

Generative programming allows tackling software problems in groups, generating code automatically for each instance of a problem. Generative programming produces these implementations from code patterns e components, following precise directive expressed through metadata, templates or domain specific languages.

I have used techniques of generative programming both in producing libraries of specializable components and to support Web applications and in particular *staged computations*.

Template Metaprogramming

Within C++, generative programming can be achieved exploiting the *template* mechanism. By means of *template metaprogramming* the compiler can be forced to perform computations on types at compile time. The mechanism is quite cumbersome but so powerful to be *Turing complete*.

Through *template metaprogramming* we have built support for *reflection* in C++ [28]. The compiler is instructed to produce a metaclass object to represent each class in a program. Such objects are used at run time to perform various tasks, e.g. serializing an object to a stream or inserting it into a persistent relational table.

Techniques of metaprogramming and reflection have been used in the development of IXE.

Code Generators and Staging

In order to provide the expressive power of template metaprogramming in a language independent way avoiding the need to resort to the meta-language, we developed a framework for code generation that allows programs to manipulate and generate code at the source level while the joining and splicing of executable code is carried out automatically at the intermediate code level. The framework introduces a type of objects to represent code fragments, and code generation is achieved assembling such code fragments like assembling construction bricks.

CodeBricks is a library implementing the framework for the .Net Common Language Runtime. The framework can be exploited by language designers to implement metaprogramming, multistage programming and *Domain Specific Languages*.

Differently from other approaches to metaprogramming that operate at the source level and require a language processor (interpreter or compiler) to be available at run time to run the resulting program, our approach operates directly on precompiled executable versions of programs.

Programs can be specialized more than once, for instance as more information becomes available to the program. Since no language processor is involved, specialization can happen at different stages through the lifetime of a program, even beyond the program build: for instance at installation time, at load time, or on the client side in a multi-tier application.

Assuming that each stage is capable of running code for the same VM, the framework can be used as support for true *multistage programming*. Manipulating code objects allows dynamically producing the various versions of a program to be executed at each stage. By exploiting transformations at the intermediate language level, computations can be expressed that are carried out on different *language processors*, not just within a single computation as allowed by traditional approaches based on source-to-source transformations (e.g. MetaML).

Research in this area has been funded in part by a grant from Microsoft Research.

Computer Graphics

On behalf of ST Microelectronics we have developed a 3D graphics library. The library implements a subset of OpenGL, and will be the base for hardware support for 3D graphics to be embedded in next generation processors for cell phones and handheld devices.

Patent Applications

US patent 20080221870 “System and method for revising natural language parse trees”, application submitted September 2008,
<http://www.freepatentsonline.com/y2008/0221870.html>

Technology Demonstrators

The following demonstrators have been developed to illustrate potential uses of the technologies I have developed:

- Deep Search (<http://semawiki.di.unipi.it/search/demo.html>), an example of semantic search on the Italian Wikipedia corpus, annotated with the tools of the TanI linguistic pipeline. Search is based on an enriched index which contains grammatical and semantic relations, so that queries can be formulated constraining names to be, for instance the subjects or objects of a given verb.
- Question Answering on Alzheimer’s disease (<http://semawiki.di.unipi.it/alzheimer/>), top performing system in Reading Comprehension Test at CLEF 2012 Pilot Task on Alzheimer’s disease.
- DeSR Parser: Italian (<http://paleo.di.unipi.it/it/parse>), Spanish (<http://paleo.di.unipi.it/es/parse>). An interactive demo of the parser in two languages.
- Yahoo! Quest (<http://quest.sandbox.yahoo.net/>), is a tool for helping users identify suitable formulations for questions to be found in the Yahoo! Answers collection. Linguistic relations are used to suggest verbs or nouns
- Yahoo! Correlator, a tool for browsing the English Wikipedia across related concepts, displaying entities, locations and persons in a visual fashion, as well as providing a summary page built from a number of relevant Wikipedia pages.

Software

As a software architect and developer I have built following tools:

- | | |
|-------------|---|
| DeSR | a deterministic incremental multilingual dependency parser, capable of handling over 20,000 words per second. Over 1200 downloads.
Available at: http://desr.sourceforge.net |
| ECL | Embeddable Common Lisp, Implementation of the Common Lisp language, embeddable in C applications. Over 21,000 downloads.
Available at: http://ecls.sourceforge.net/ . |
| CMM | customizable garbage collector for C++, used by Sun Microsystems in the implementation of Java.
Available at: http://www.di.unipi.it/~attardi/cmm.html . |
| IXE | high performance C++ class library for indexing and search based on template metaprogramming techniques.
http://medialab.di.unipi.it/Project/IXE/doc |
| IXE Crawler | parallel Web crawler written in C++, achieving high throughput by means of both multithreading and asynchronous IO |
| NeTagger | a Named Entity tagger based on a Maximum Entropy classifier. |
| SRL | a Semantic Role Labeler, based on multiple learning classifiers, used at the CoNLL 2008 Shared task. |

I also contributed to various software utilities: TreeTagger, cURL, rddtool, libssh2.

The following are software projects developed with the students of the course LOA:

LoaCMS a Content Management System, available at <http://software.freshmeat.net/~loa/>.

Loacker an e-learning platform, available at <http://sourceforge.net/projects/loacker/>.

LoaMA a Personal Media Aggregator, available at <http://sourceforge.net/projects/loama/>.

GALOAP Game with a Purpose for annotating dependency trees for sentences (<http://galoap.codeplex.com/>). Winner of the INSEMTIVES Game Idea Challenge 2010 (<http://challenge.insemtives.eu/>).

Selected Publications

Journals

- [1] G. Attardi, M. Simi. A Description Oriented Logic for Building Knowledge Bases. *Proceedings of the IEEE*, **74**(10), 1335–1344, 1986.
- [2] G. Attardi. The Embeddable Common Lisp, *ACM Lisp Pointers*, 8(1), 30–41, 1995. <http://ecls.sourceforge.net>.
- [3] G. Attardi, C. Traverso. Strategy-accurate parallel Buchberger algorithms, *Journal of Symbolic Computing*, **22**, 1–15, 1996.
- [4] G. Attardi, M. Gaspari. Multilanguage Interoperability. *Computers and Artificial Intelligence*, **15**(6), 531–554, 1996.
- [5] G. Attardi, M. Simi. Communication across Viewpoints, *Journal of Logic, Language and Information*, **7**, 53–75, 1998.
- [6] G. Attardi, T. Flagella, P. Iglio. A customisable memory management framework for C++, *Software: Practice and Experience*, **28**(11), 1143–1183, 1998.
- [7] G. Attardi, A. Cisternino, M. Simi. Web-based Configuration Assistants, *Artificial Intelligence for Engineering Design, Analysis and Manufacturing*, **12**(3), 321–331, 1998.
- [8] G. Attardi, S. Di Marco, D. Salvi. Categorisation by context, *Journal of Universal Computer Science*, **4**(9), 1998.
- [9] G. Attardi, A. Cisternino, A. Kennedy. "CodeBricks: Code Fragments as Building Blocks. *ACM SIGPLAN Notices*, **38**(10), 66–74, October, 2003.
- [10] G. Attardi, A. Cisternino. "Multistage programming support in CLI. *IEE Proceedings Software*, 275–282, **150**(5), October, 2003.
- [11] G. Attardi, A. Cisternino, D. Colombo. "CIL + Metadata > Executable. *Journal of Object Technology*, **3**(2), March-April, 2004.

Conferences

- [12] G. Attardi, C. Montangero and G. Prini, A High Level Machine for Artificial Intelligence, *Proc. of AISB Summer Conference*, Edinburgh, 1976.
- [13] G. Attardi, M. Simi. Consistency and Completeness of Omega, a Logic for Knowledge Representation. *Proc. of Seventh International Joint Conference on Artificial Intelligence*, Vancouver, August 1981, 504–510.
- [14] G. Attardi, M. Simi. Extending the Power of Programming by Examples. *Proceedings of SIGOA Conference on Office Information Systems*, giugno 1982, *SIGOA Newsletter* vol. 3, n. 1 e 2. Also in *Integrated Interactive Computing Systems*, P. Degano, E. Sandewall (Editors), North-Holland, 1983.
- [15] G. Attardi, M. Simi. Metalanguage and Reasoning Across Viewpoints. *Proc. of the Sixth European Conference on Artificial Intelligence*, T. O'Shea (Editor), North Holland, 1984.
- [16] G. Attardi et al. Taxonomic Reasoning. *Proc. of the 7th European Conference on Artificial Intelligence*, Brighton, 1986, 130–139, also in *Advances in Artificial Intelligence II*, B. Du Boulay, D. Hogg and L. Steels (Editors), North-Holland, 1987, 277–286.
- [17] G. Attardi et al. Towards the development of a new standard for LISP. *Proc. of the 7th European Conference on Artificial Intelligence*, Brighton, *Advances in Artificial Intelligence II*, B. Du Boulay, D. Hogg and L. Steels (Editors), North-Holland, 1987.
- [18] G. Attardi, I. Filotti, J. Marks. Techniques for Dynamic Software Migration. *Proc. of the 5th Annual ESPRIT Conference*, CEC (Editors), North Holland, 1988, 475–491.
- [19] G. Attardi et al. Metalevel Programming in CLOS. *Proc. of the Third ECOOP*, Cambridge University Press, 1989.
- [20] G. Attardi, M. Simi. Reflections about reflection. *Proc. of 2nd Int. Conference on Principles of Knowledge Representation and Reasoning*, Morgan-Kaufmann, 1991.

- [21] G. Attardi, M. Simi. Proofs in context, in Doyle, J. and Torasso, P. (eds.) *Principles of Knowledge Representation and Reasoning: Proc. of the Fourth International Conference*, Morgan Kaufmann, San Mateo, California, 1994.
- [22] G. Attardi, C. Traverso. The PoSSo Library for Polynomial System Solving, *Proc. of AIHENP95*, World Scientific Publishing Company, Singapore, 1995.
- [23] G. Attardi, S. Di Marco, D. Salvi. Categorisation by context, *Best Full Paper Award, Proc. of WebNet 1998*, Orlando, Florida, 1998.
- [24] G. Attardi, A. Gulli, F. Sebastiani. Theseus: Categorization by context, *8th World Wide Web Conference*, Toronto, Canada, 1999.
- [25] G. Attardi, A. Cisternino, [Reflection support by means of template metaprogramming](#), *Proc. of Third International Conference on Generative and Component-Based Software Engineering, LNCS 2186*, 178-187, Springer-Verlag, Berlin, 2001.
- [26] G. Attardi, A. Cisternino, [Template Metaprogramming an Object Interface to Relational Tables](#), *Reflection 2001, LNCS 2192*, 266-267, Springer-Verlag, Berlin, 2001.
- [27] G. Attardi, A. Cisternino, F. Formica, M. Simi, A. Tommasi, C. Zavattari. [PIQASso: Pisa Question Answering System](#). *Proc. of Text Retrieval Conference (Trec-10)*, 599–607, NIST, Gaithersburg (MD), November 13–16, 2001.
- [28] G. Attardi, A. Cisternino. Self Reflection for Adaptive Programming. *Generative Programming and Component Engineering 2002*, Don S. Batory, Charles Consel, Walid Taha (eds), Pittsburgh (PA), LNCS 2487, 50–65, Springer-Verlag, Berlin, 2002.
- [29] G. Attardi, A. Cisternino, F. Formica, M. Simi, A. Tommasi. Web suggestions and robust validation for QA. In *Proc. of Text Retrieval Conference 2002*, NIST, Gaithersburg (MD), 2002.
- [30] G. Attardi, A. Cisternino, A. Kennedy. CodeBricks: Code Fragments as Building Blocks. *Proc. of the 2003 ACM SIGPLAN Workshop on Partial Evaluation and Semantics Based Program Manipulation*, S. Diego (CA), 2003. Also in *ACM SIGPLAN Notices*, Vol. 38, N. 10, 66-74, October, 2003.
- [31] G. Attardi, A. Esuli, C. Patel. Using Clustering and Blade Clusters in the TeraByte task. *Proc. of Text Retrieval Conference (Trec-13)*, NIST, Gaithersburg (MD), 2004.
- [32] G. Attardi, A. Esuli, M. Simi, "Best bets: thousands of queries in search of a client", *13th World Wide Web Conference*, New York, USA, pp. 422-423, 2004.
- [33] G. Attardi. IXE at the TREC 2005 TeraByte task. *Proc. of Text Retrieval Conference (Trec-14)*, NIST, Gaithersburg (MD), 2005.
- [34] G. Attardi. Experiments with a Multilanguage non-projective dependency parser. In *Proc. of CoNLL-X Shared Task*, 2006.
- [35] G. Attardi, M. Simi. Extracting Dependency Relations for Opinion Mining, In *Proc. of Workshop on Distributed Agent-based Retrieval Tools*, Cagliari, 2006.
- [36] G. Attardi, M. Simi. Blog Mining through Opinionated Words, In *Proc. of TREC-15*, NIST, Gaithersburg, 2006.
- [37] G. Attardi and M. Ciaramita. 2007. Tree Revision Learning for Dependency Parsing. In *Proc. of HLT/ACL, 2007*.
- [38] G. Attardi and M. Ciaramita. 2007. Multilingual Dependency Parsing and Domain Adaptation with DeSR. CoNLL 2007 Shared Task.
- [39] G. Attardi . 2007. A Framework for Incremental Non-Projective Dependency Parsing. Submitted.
- [40] M. Ciaramita and G. Attardi. 2007. Dependency Parsing with Second-Order Feature Maps and Annotated Semantic Information. *Proc. of the 10th International Conference on Parsing Technologies (IWPT)*.
- [41] G. Attardi, M. Ciaramita. 2007. Tree revision learning for dependency parsing, Proc. NAACL HLT 2007, 388-395, Rochester.
- [42] G. Attardi, M. Simi, F. Dell'orletta, A. Chanev, M. Ciaramita. 2007. Multilingual Dependency Parsing and Domain Adaptation using DeSR. *Proc of CoNLL Shared Task Session of of EMNLP-CoNLL 2007*, 1112-1118, Prague.
- [43] H. Zaragoza, et al. 2007. Ranking Very Many Typed Entities on Wikipedia. *Proc. of CIKM 2007*, Lisboa.
- [44] G. Attardi, M. Simi. 2007. DeSR at the Evalita Dependency Parsing Task *Proc. of Workshop Evalita 2007. Intelligenza Artificiale*, 4(2).

- [45] H. Zaragoza, J. Atserias, M. Ciaramita, and G. Attardi. 2008. Semantically annotated snapshot of the English Wikipedia. *Proc. LREC 2008*.
- [46] G. Attardi, F. Dell'Orletta. 2008. Chunking and Dependency Parsing. *Proc. of Workshop on Partial Parsing*, LREC 2008.
- [47] C. Bosco, A. Mazzei, V. Lombardo, G. Attardi, A. Corazza, A. Lavelli, L. Lesmo, G. Satta, M. Simi. Comparing Italian parsers on a common treebank: the Evalita experience *Proceedings of LREC 2008*, Marrakech, 2008.
- [48] G. Attardi, F. Dell'Orletta. Reverse Revision and Linear Tree Combination for Dependency Parsing. *Proc. of NAACL HLT 2009*, 2009.
- [49] G. Attardi, M. Simi. Overview of the EVALITA 2009 Part-of-Speech Tagging Task *Proc. of Workshop Evalita 2009*, ISBN 978-88-903581-1-1, 2009.
- [50] G. Attardi, F. Dell'Orletta, M. Simi, J. Turian. Accurate Dependency Parsing with a Stacked Multilayer Perceptron. *Proc. of Workshop Evalita 2009*, ISBN 978-88-903581-1-1, 2009.
- [51] G. Attardi, S. Dei Rossi, F. Dell'Orletta, E.M. Vecchi. The Tan1 Named Entity Recognizer at Evalita 2009. *Proc. of Workshop Evalita 2009*, ISBN 978-88-903581-1-1, 2009.
- [52] G. Attardi. Text Analytics for Wikipedia Semantic Search. *4th Workshop on the Future of Web Search: Semantic Search*, Ibiza, Spain, April 17-18, 2009.
- [53] G. Attardi, S. Dei Rossi, F. Dell'Orletta, E.M. Vecchi. Experiments in tagger combination: arbitrating, guessing, correcting, suggesting. *Proc. of Workshop Evalita 2009*, ISBN 978-88-903581-1-1, 2009.
- [54] J. Atserias, G. Attardi, M. Simi, H. Zaragoza. Active Learning for Building a Corpus of Questions for Parsing. *Proc. of LREC 2010*, 2010.
- [55] G. Attardi, S. Dei Rossi, G. Di Pietro, A. Lenci, S. Montemagni, M. Simi. A Resource and Tool for Super-sense Tagging of Italian Texts. *Proc. of LREC 2010*, 2010.
- [56] C. Bosco, S. Montemagni, A. Mazzei, V. Lombardo, F. Dell'Orletta, A. Lenci, L. Lesmo, G. Attardi, M. Simi, A. Lavelli, J. Hall, J. Nilsson and J. Nivre. Comparing the influence of different treebank annotations on dependency parsing performance. *Proc. of LREC 2010*, Malta, 2010.
- [57] G. Attardi, S. Dei Rossi, M. Simi. The Tan1 Pipeline. *Proc. of LREC Workshop on WSPP*, Malta, 2010.
- [58] G. Attardi, S. Dei Rossi, M. Simi. Tan1-1: Coreference Resolution by Parse Analysis and Similarity Clustering. *Proc. of SemEval 2010*, Uppsala, 2010.
- [59] G. Attardi, S. Dei Rossi, M. Simi. Dependency Parsing of Indian Languages with DeSR. *Proc. of ICON-2010 tools contest on Indian language dependency parsing*, Kharagpur, India, 2010.
- [60] G. Attardi, A. Chanev, A.V. Miceli Barone. A Dependency Based Statistical Translation Model. *Proc. of Fifth Workshop on Syntax, Semantics and Structure in Statistical Translation*, ACL HLT 2011, Portland, 2011.
- [61] G. Attardi. [Internet è di Tutti](#). Conferenza GARR 2011, Bologna, 8/11/2011.
- [62] M. Ciaramita, G. Attardi. Dependency parsing with second-order feature maps and annotated semantic information. In H. Bunt, P. Merlo and J. Nivre (eds.), [Trends in Parsing Technology](#). Text, Speech and Language Technology Series, Vol 43, pp. 87-104. Springer. 2011.
- [63] G. Attardi, D. Sartiano, M. Simi. Active Learning for Domain Adaptation of Dependency Parsing on Legal Texts. *Proc. of Workshop on Semantic Processing of Legal Texts (SPLeT of LREC-2012)*, Istanbul, Turchia, May 2012.
- [64] G. Attardi, S. Dei Rossi, M. Simi. The Evalita 2011 Anaphora Resolution Task: assessment without comparison. *Proc. of Evalita 2011*. Springer LNCS (to appear), 2012.
- [65] G. Attardi, M. Simi, A. Zanelli. Domain Adaptation by Active Learning. *Proc. of Evalita 2011*, Springer LNCS (to appear), 2012.
- [66] G. Attardi, S. Dei Rossi, M. Simi. The Tan1 Lemmatizer Enriched with a Sequence of Cascading Filters. *Proc. of Evalita 2011*, Springer LNCS (to appear), 2012.
- [67] G. Attardi, G. Berardi, S. Dei Rossi, M. Simi. The Tan1 Tagger for Named Entity Recognition on Transcribed Broadcast News at Evalita 2011. *Proc. of Evalita 2011*, Springer LNCS (to appear), 2012.
- [68] G. Attardi, M. Simi, A. Zanelli. Tuning DeSR for Dependency Parsing of Italian. *Proc. of Evalita 2011*, Springer LNCS (to appear), 2012.
- [69] G. Attardi, L. Baronti, S. Dei Rossi, M. Simi. SuperSense Tagging with a Maximum Entropy Markov Model. *Proc. of Evalita 2011*, Springer LNCS (to appear), 2012.

References

Ricardo Baeza-Yates (rby@yahoo-inc.com), Yahoo! Research.

Jeffrey Ullman, (ullman@cs.stanford.edu), Stanford University.

Luca Cardelli (luca@microsoft.com), Microsoft Research, Cambridge, UK.