|  |
| --- |
| Nr 2 |
| Tipologia prodotto Contributo in Collana |
| Elenco autori Monica Monachini, Valeria Quochi |
| Titolo Tour de CLARIN: Italy |
| Codice identificativo (ISBN) 9789082990911 |
| anno pubblicazione 2019 |
| Altre informazioni Tour de CLARIN Volume Two url: https://office.clarin.eu/v/CE-2019-1537-Tour-de-CLARIN-volume-II-2019.pdf doi: 10.5281/zenodo.3754164 |

|  |
| --- |
| Nr 3 |
| Tipologia prodotto Contributo in Conferenza o Convegno |
| Elenco autori Lionel Nicolas, Alexander König, Monica Monachini, Riccardo Del Gratta, Silvia Calamai, Andrea Abel, Alessandro Enea, Francesca Biliotti, Valeria Quochi, Francesco Vincenzo Stella |
| Titolo CLARIN-IT: State of Affairs, Challenges and Opportunities |
| Codice identificativo (ISBN) 978-91-7685-273-6 |
| anno pubblicazione 2018 |
| abstract: his paper gives an overview on the Italian national CLARIN consortium as it currently stands two years after its creation at the end of 2015. It thus discusses the current state of affairs of the consortium on several aspects, especially with regards to members. It also discusses the events and initiatives that have been undertaken, as well as the ones that are planned in the close future. It finally outlines the conclusions of a user survey performed to understand the expectations of a targeted user population and provides indications regarding the next steps planned.Selected papers from the CLARIN Annual Conference 2017, Budapest, 18-20 September 2017 url: http://www.ep.liu.se/ecp/contents.asp?issue=147 abstract: his paper gives an overview on the Italian national CLARIN consortium as it currently stands two years after its creation at the end of 2015. It thus discusses the current state of affairs of the consortium on several aspects, especially with regards to members. It also discusses the events and initiatives that have been undertaken, as well as the ones that are planned in the close future. It finally outlines the conclusions of a user survey performed to understand the expectations of a targeted user population and provides indications regarding the next steps planned. |

|  |
| --- |
| Nr 4 |
| Tipologia prodotto Contributo in Conferenza o Convegno (Senza ISBN) |
| Titolo CLARIN-IT: State of Affairs, Challenges and Opportunities |
| Descrizione |
| Elenco autori Lionel Nicolas, Alexander Konig, Monica Monachini, Riccardo Del Gratta, Silvia Calamai, Andrea Abel, Alessandro Enea, Francesca Biliotti, Valeria Quochi |
| Ruolo svolto |
| anno pubblicazione 2017 |
| abstract: This paper provides an overview on the Italian national CLARIN consortium and the status of CLARIN-IT in general. It thus discusses the current state of affairs of the consortium and provi-des information on the members, especially with regards to what they offer to CLARIN in terms of resources, services and expertise, and what CLARIN offers them to further their own research. url: https://www.clarin.eu/event/2017/clarin-annual-conference-2017-budapest-hungary abstract: This paper provides an overview on the Italian national CLARIN consortium and the status of CLARIN-IT in general. It thus discusses the current state of affairs of the consortium and provi-des information on the members, especially with regards to what they offer to CLARIN in terms of resources, services and expertise, and what CLARIN offers them to further their own research. |

|  |
| --- |
| Nr 6 |
| Tipologia prodotto Contributo in Collana |
| Elenco autori Valeria Quochi |
| Titolo Development and representation of Italian light-fare constructions |
| Codice identificativo (ISBN) 9789027204417 |
| anno pubblicazione 2016 |
| Altre informazioni Corpus-based Approaches to Construction Grammar url: https://benjamins.com/#catalog/books/cal.19.03quo/details doi: 10.1075/cal.19.03quo ISSN: 1573-594X |

|  |
| --- |
| Nr 7 |
| Tipologia prodotto Contributo in Conferenza o Convegno (Senza ISBN) |
| Titolo Generative Lexicon and polysemy: inducing logical alternations |
| Descrizione |
| Elenco autori Francesca Frontini, Valeria Quochi, Monica Monachini |
| Ruolo svolto |
| anno pubblicazione 2015 |
| abstract: The current paper brings together the results of a series of experiments for inducing regular sense alternations, or regular/ logical polysemy, from a computational lexicon based on the Generative Lexicon theory. The results are discussed in light of the potential benefits and uses of the amended algorithm. abstract: The current paper brings together the results of a series of experiments for inducing regular sense alternations, or regular/ logical polysemy, from a computational lexicon based on the Generative Lexicon theory. The results are discussed in light of the potential benefits and uses of the amended algorithm. |

|  |
| --- |
| Nr 8 |
| Tipologia prodotto Contributo in Conferenza o Convegno |
| Elenco autori De Felice I., Bartolini R., Russo I., Quochi V., Monachini M. |
| Titolo Evaluating ImagAct-WordNet mapping for English and Italian through videos |
| Codice identificativo (ISBN) 978-88-67-41472-7 |
| anno pubblicazione 2014 |
| abstract: In this paper we present the results of the evaluation of an automatic mapping between two lexical resources, WordNet/ItalWordNet and ImagAct, a conceptual ontology of action types instantiated by video scenes. Results are compared with those obtained from a previous experiment performed only on Italian data. Differences between the two evaluation strategies, as well as between the quality of the mappings for the two languages considered in this paper, are iscussed. url: http://clic.humnet.unipi.it/proceedings/Proceedings-CLICit-2014.pdf abstract: In this paper we present the results of the evaluation of an automatic mapping between two lexical resources, WordNet/ItalWordNet and ImagAct, a conceptual ontology of action types instantiated by video scenes. Results are compared with those obtained from a previous experiment performed only on Italian data. Differences between the two evaluation strategies, as well as between the quality of the mappings for the two languages considered in this paper, are iscussed. |

|  |
| --- |
| Nr 9 |
| Tipologia prodotto Contributo in Conferenza o Convegno |
| Elenco autori Bartolini R., Quochi V., De Felice, I., Russo I., Monachini M. |
| Titolo From Synsets to Videos: Enriching ItalWordNet Multimodally |
| Codice identificativo (ISBN) 978-2-9517408-8-4 |
| anno pubblicazione 2014 |
| abstract: The paper describes the multimodal enrichment of ItalWordNet action verbs' entries by means of an automatic mapping with a conceptual ontology of action types instantiated by video scenes (ImagAct). The two resources present significative differences as well as interesting complementary features, such that a mapping of these two resources can lead to a an enrichment of IWN, through the connection between synsets and videos apt to illustrate the meaning described by glosses. Here, we describe an approach inspired by ontology matching methods for the automatic mapping of ImagAct video scenes onto ItalWordNet. The experiments described in the paper are conducted on Italian, but the same methodology can be extended to other languages for which WordNets have been created, since ImagAct is available also for English, Chinese and Spanish. This source of multimodal information can be exploited to design second language learning tools, as well as for language grounding in action recognition in video sources and potentially for robotics. url: http://www.lrec-conf.org/proceedings/lrec2014/index.html abstract: The paper describes the multimodal enrichment of ItalWordNet action verbs\' entries by means of an automatic mapping with a conceptual ontology of action types instantiated by video scenes (ImagAct). The two resources present significative differences as well as interesting complementary features, such that a mapping of these two resources can lead to a an enrichment of IWN, through the connection between synsets and videos apt to illustrate the meaning described by glosses. Here, we describe an approach inspired by ontology matching methods for the automatic mapping of ImagAct video scenes onto ItalWordNet. The experiments described in the paper are conducted on Italian, but the same methodology can be extended to other languages for which WordNets have been created, since ImagAct is available also for English, Chinese and Spanish. This source of multimodal information can be exploited to design second language learning tools, as well as for language grounding in action recognition in video sources and potentially for robotics. |

|  |
| --- |
| Nr 10 |
| Tipologia prodotto Contributo in Conferenza o Convegno |
| Elenco autori Antico Giovanni, Valeria Quochi, Monica Monachini, Maurizio Martinelli |
| Titolo Marrying Technical Writing with LRT |
| Codice identificativo (ISBN) 978-2-9517408-8-4 |
| anno pubblicazione 2014 |
| abstract: In the last years the Technical Writer operational scenarios and the workflow sensibly changed; specifically,"free style" writing - or manual writing - has become outdated and technical writing is now much more concerned with structured management of content than in the past. Technical writing has become more demanding due to a number of factors among which the rise and spread of mobile devices usage. This paper discusses the new needs of technical writing and content management business and how LRT can help it improve quality and productivity. url: http://www.lrec-conf.org/proceedings/lrec2014/index.html abstract: In the last years the Technical Writer operational scenarios and the workflow sensibly changed; specifically,\"free style\" writing - or manual writing - has become outdated and technical writing is now much more concerned with structured management of content than in the past. Technical writing has become more demanding due to a number of factors among which the rise and spread of mobile devices usage. This paper discusses the new needs of technical writing and content management business and how LRT can help it improve quality and productivity. |

|  |
| --- |
| Nr 11 |
| Tipologia prodotto Contributo in Conferenza o Convegno |
| Elenco autori Frontini F., Quochi V., Monachini M. |
| Titolo Polysemy alternations extraction using the PAROLE SIMPLE CLIPS Italian lexicon |
| Codice identificativo (ISBN) 978-88-67-41472-7 |
| anno pubblicazione 2014 |
| abstract: This paper presents the results of an experiment of polysemy alternations induction from a lexicon (Utt and Pad´o, 2011; Frontini et al., 2014), discussing the results and proposing an amendment in the original algorithm. url: http://clic.humnet.unipi.it/proceedings/Proceedings-CLICit-2014.pdf abstract: This paper presents the results of an experiment of polysemy alternations induction from a lexicon (Utt and Pad´o, 2011; Frontini et al., 2014), discussing the results and proposing an amendment in the original algorithm. |

|  |
| --- |
| Nr 12 |
| Tipologia prodotto Contributo in Conferenza o Convegno |
| Elenco autori Frontini Francesca, Valeria Quochi, Sebastian Padó, Jason Utt, Monica Monachini |
| Titolo Polysemy Index for Nouns: an Experiment on Italian using the PAROLE SIMPLE CLIPS Lexical Database |
| Codice identificativo (ISBN) 978-2-9517408-8-4 |
| anno pubblicazione 2014 |
| abstract: An experiment is presented to induce a set of polysemous basic type alternations (such as ANIMAL-FOOD, or BUILDING-INSTITUTION) by deriving them from the sense alternations found in an existing lexical resource. The paper builds on previous work and applies those results to the Italian lexicon PAROLE SIMPLE CLIPS. The new results show how the set of frequent type alternations that can be induced from the lexicon is partly different from the set of polysemy relations selected and explicitly applied by lexicographers when building it. The analysis of mismatches shows that frequent type alternations do not always correspond to prototypical polysemy relations, nevertheless the proposed methodology represents a useful tool offered to lexicographers to systematically check for possible gaps in their resource.LREC 2014 Ninth International Conference on Language Resources and Evaluation Proceedings url: http://www.lrec-conf.org/proceedings/lrec2014/index.html abstract: An experiment is presented to induce a set of polysemous basic type alternations (such as ANIMAL-FOOD, or BUILDING-INSTITUTION) by deriving them from the sense alternations found in an existing lexical resource. The paper builds on previous work and applies those results to the Italian lexicon PAROLE SIMPLE CLIPS. The new results show how the set of frequent type alternations that can be induced from the lexicon is partly different from the set of polysemy relations selected and explicitly applied by lexicographers when building it. The analysis of mismatches shows that frequent type alternations do not always correspond to prototypical polysemy relations, nevertheless the proposed methodology represents a useful tool offered to lexicographers to systematically check for possible gaps in their resource. |

|  |
| --- |
| Nr 13 |
| Tipologia prodotto Contributo in Conferenza o Convegno |
| Elenco autori Alessandro Panunzi, Irene De Felice, Lorenzo Gregori, Stefano Jacoviello, Monica Monachini, Massimo Moneglia, Valeria Quochi, Irene Russo |
| Titolo Translating action verbs using a dictionary of images: the IMAGACT ontology. |
| Codice identificativo (ISBN) 978-88-88906-97-3 |
| anno pubblicazione 2014 |
| abstract: Action verbs have many meanings, covering actions in different ontological types. Moreover, each language categorizes action in its own way. One verb can refer to many different actions and one action can be identified by more than one verb. The range of variations within and across languages is largely unknown, causing trouble in all translation tasks. IMAGACT is a corpus-based ontology of action concepts, derived from English and Italian spontaneous speech corpora, which makes use of the universal language of images to identify the different action types extended by verbs referring to action in English, Italian, Chinese and Spanish. This paper presents the IMAGACT search interface and the various kinds of linguistic information the user can derive from it. IMAGACT makes explicit the variation of meaning of action verbs within one language and allows comparisons of verb variations within and across languages. Because the action concepts are represented with videos, extension into new languages beyond those presently implemented in IMAGACT is done using competence-based judgments by mother-tongue informants, without intense lexicographic work involving underdetermined semantic descriptions.Proceedings of the XVI EURALEX International Congress: The User in Focus. url: http://euralex2014.eurac.edu/en/callforpapers/Documents/EURALEX%202014\_gesamt.pdf abstract: Action verbs have many meanings, covering actions in different ontological types. Moreover, each language categorizes action in its own way. One verb can refer to many different actions and one action can be identified by more than one verb. The range of variations within and across languages is largely unknown, causing trouble in all translation tasks. IMAGACT is a corpus-based ontology of action concepts, derived from English and Italian spontaneous speech corpora, which makes use of the universal language of images to identify the different action types extended by verbs referring to action in English, Italian, Chinese and Spanish. This paper presents the IMAGACT search interface and the various kinds of linguistic information the user can derive from it. IMAGACT makes explicit the variation of meaning of action verbs within one language and allows comparisons of verb variations within and across languages. Because the action concepts are represented with videos, extension into new languages beyond those presently implemented in IMAGACT is done using competence-based judgments by mother-tongue informants, without intense lexicographic work involving underdetermined semantic descriptions. |

|  |
| --- |
| Nr 14 |
| Tipologia prodotto Contributo in Conferenza o Convegno |
| Elenco autori Fahad Khan, Francesca Frontini, Riccardo Del Gratta, Monica Monachini, Valeria Quochi |
| Titolo Generative Lexicon Theory and Linguistic Linked Open Data |
| Codice identificativo (ISBN) 978-1-937284-98-5 |
| anno pubblicazione 2013 |
| abstract: In this paper we look at how Generative Lexicon theory can assist in providing a more thorough definition of word senses as links between items in a RDF-based lexicon and concepts in an ontology. We focus on the definition of lexical sense in lemon and show its limitations before defining a new model based on lemon and which we term lemonGL. This new model is an initial attempt at providing a way of structuring lexico-ontological resources as linked data in such a way as to allow a rich representation of word meaning (following the GL theory) while at the same time (attempting to) re-main faithful to the separation between the lexicon and the ontology as recommended by the lemon model.Proceedings of the 6th International Conference on Generative Approaches to the Lexicon. Generative Lexicon and Distributional Semantics. url: http://aclweb.org/anthology/W13-54 abstract: In this paper we look at how Generative Lexicon theory can assist in providing a more thorough definition of word senses as links between items in a RDF-based lexicon and concepts in an ontology. We focus on the definition of lexical sense in lemon and show its limitations before defining a new model based on lemon and which we term lemonGL. This new model is an initial attempt at providing a way of structuring lexico-ontological resources as linked data in such a way as to allow a rich representation of word meaning (following the GL theory) while at the same time (attempting to) re-main faithful to the separation between the lexicon and the ontology as recommended by the lemon model. |

|  |
| --- |
| Nr 15 |
| Tipologia prodotto Contributo in Conferenza o Convegno |
| Elenco autori Quochi, Valeria; Frontini, Francesca; Rubino, Francesco |
| Titolo A MWE Acquisition and Lexicon Builder Web Service |
| Codice identificativo (ISBN) 9781627483896 |
| anno pubblicazione 2012 |
| abstract: This paper describes the development of a web-service tool for the automatic extraction of Multi-word expressions lexicons, which has been integrated in a distributed platform for the automatic creation of linguistic resources. The main purpose of the work described is thus to provide a (computationally "light") tool that produces a full lexical resource: multi-word terms/items with relevant and useful attached information that can be used for more complex processing tasks and applications (e.g. parsing, MT, IE, query expansion, etc.). The output of our tool is a MW lexicon formatted and encoded in XML according to the Lexical Mark-up Framework. The tool is already functional and available as a service. Evaluation experiments show that the tool precision is of about 80\%.Proceedings of COLING 2012: Technical Papers url: http://aclweb.org/anthology/C/C12/C12-1140.pdf abstract: This paper describes the development of a web-service tool for the automatic extraction of Multi-word expressions lexicons, which has been integrated in a distributed platform for the automatic creation of linguistic resources. The main purpose of the work described is thus to provide a (computationally \"light\") tool that produces a full lexical resource: multi-word terms/items with relevant and useful attached information that can be used for more complex processing tasks and applications (e.g. parsing, MT, IE, query expansion, etc.). The output of our tool is a MW lexicon formatted and encoded in XML according to the Lexical Mark-up Framework. The tool is already functional and available as a service. Evaluation experiments show that the tool precision is of about 80\%. |

|  |
| --- |
| Nr 16 |
| Tipologia prodotto Contributo in Conferenza o Convegno |
| Elenco autori Francesca Frontini, Valeria Quochi, Francesco Rubino |
| Titolo Automatic Creation of Quality Multi-Word Lexica from Noisy Text Data |
| Codice identificativo (ISBN) 978-1-4503-1919-5 |
| anno pubblicazione 2012 |
| abstract: This paper describes the design of a tool for the automatic creation of multi-word lexica that is deployed as a web service and runs on automatically web-crawled data within the framework of the PANACEA platform. The main purpose of our task is to provide a (computationally "light") tool that creates a full high quality lexical resource of multi-word items. Within the platform, this tool is typically inserted in a work flow whose first step is automatic web-crawling. Therefore, the input data of our lexical extractor is intrinsically noisy. The paper evaluates the capacity of the tool to deal with noisy data, and in particular with texts containing a significant amount of duplicated paragraphs. The accuracy of the extraction of multi-word expressions from the original crawled corpus is compared to the accuracy of the extraction from a later "de-duplicated" version of the corpus. The paper shows how our method can extract with sufficiently good precision also from the original, noisy crawled data. The output of our tool is a multi-word lexicon formatted and encoded in XML according to the Lexical Mark-up Framework.Proceedings of the Sixth Workshop on Analytics for Noisy Unstructured Text Data url: http://www.kde.cs.tut.ac.jp/~aono/pdf/COLING2012/AND/pdf/AND04.pdf abstract: This paper describes the design of a tool for the automatic creation of multi-word lexica that is deployed as a web service and runs on automatically web-crawled data within the framework of the PANACEA platform. The main purpose of our task is to provide a (computationally \"light\") tool that creates a full high quality lexical resource of multi-word items. Within the platform, this tool is typically inserted in a work flow whose first step is automatic web-crawling. Therefore, the input data of our lexical extractor is intrinsically noisy. The paper evaluates the capacity of the tool to deal with noisy data, and in particular with texts containing a significant amount of duplicated paragraphs. The accuracy of the extraction of multi-word expressions from the original crawled corpus is compared to the accuracy of the extraction from a later \"de-duplicated\" version of the corpus. The paper shows how our method can extract with sufficiently good precision also from the original, noisy crawled data. The output of our tool is a multi-word lexicon formatted and encoded in XML according to the Lexical Mark-up Framework. |

|  |
| --- |
| Nr 17 |
| Tipologia prodotto Deliverable di Progetto |
| Titolo D3.4 Third version (v4) of the integrated platform and documentation |
| Descrizione |
| Elenco autori Poch, Marc; Hamon, Olivier; Quochi, Valeria; Del Gratta, Riccardo; Toral, Antonio; Thurmair, Gregor; Prokopidis, Prokopis; Bel, Nuria |
| Ruolo svolto |
| anno pubblicazione 2012 |
| Altre informazioni |

|  |
| --- |
| Nr 18 |
| Tipologia prodotto Deliverable di Progetto |
| Titolo D6.2 Integrated Final Version of the Components for Lexical Acquisition |
| Descrizione |
| Elenco autori Rimell, Laura; Bel, Núria; Padró, Muntsa; Frontini, Francesca; Monachini, Monica; Quochi, Valeria |
| Ruolo svolto |
| anno pubblicazione 2012 |
| Altre informazioni |

|  |
| --- |
| Nr 19 |
| Tipologia prodotto Deliverable di Progetto |
| Titolo D6.3 Monolingual lexica for English, Spanish and Italian tuned for a particular domain (LAB and ENV) |
| Descrizione |
| Elenco autori Rimell, Laura; Bel, Nuria; Padrò, Muntsa; Frontini, Francesca; Monachini, Monica; Quochi, Valeria; Del Gratta, Riccardo |
| Ruolo svolto |
| anno pubblicazione 2012 |
| Altre informazioni |

|  |
| --- |
| Nr 20 |
| Tipologia prodotto Deliverable di Progetto |
| Titolo D6.5 Merged dictionaries |
| Descrizione |
| Elenco autori Rimell, Laura; Bel, Núria; Padró, Muntsa; Frontini, Francesca; Monachini, Monica; Quochi, Valeria; Del Gratta, Riccardo |
| Ruolo svolto |
| anno pubblicazione 2012 |
| Altre informazioni |

|  |
| --- |
| Nr 21 |
| Tipologia prodotto Deliverable di Progetto |
| Titolo D7.4 Third evaluation report. Evaluation of PANACEA v3 and produced resources |
| Descrizione |
| Elenco autori Quochi, Valeria; Frontini, Francesca; Bartolini, Roberto; Hamon, Olivier; Poch Riera, Marc; Padro, Muntsa; Bel, Nuria; Thurmair, Gregor; Toral, Antonio; Kamran, Amir |
| Ruolo svolto |
| anno pubblicazione 2012 |
| Altre informazioni |

|  |
| --- |
| Nr 22 |
| Tipologia prodotto Contributo in Conferenza o Convegno |
| Elenco autori Caselli, Tommaso; Frontini, Francesca; Quochi, Valeria; Rubino, Francesco and Russo, Irene |
| Titolo Flexible Acquisition of Subcategorization Frames in Italian |
| Codice identificativo (ISBN) 9782951740877 |
| anno pubblicazione 2012 |
| abstract: Lexica of predicate-argument structures constitute a useful tool for several tasks in NLP. This paper describes a web-service system for automatic acquisition of verb subcategorization frames (SCFs) from parsed data in Italian. The system acquires SCFs in an unsupervised manner. We created two gold standards for the evaluation of the system, the first by mixing together information from two lexica (one manually created and the second automatically acquired) and manual exploration of corpus data and the other annotating data extracted from a specialized corpus (environmental domain). Data filtering is accomplished by means of the maximum likelihood estimate (MLE). The evaluation phase has allowed us to identify the best empirical MLE threshold for the creation of a lexicon (P=0.653, R=0.557, F1=0.601). In addition to this, we assigned to the extracted entries of the lexicon a confidence score based on the relative frequency and evaluated the extractor on domain specific data. The confidence score will allow the final user to easily select the entries of the lexicon in terms of their reliability: one of the most interesting feature of this work is the possibility the final users have to customize the results of the SCF extractor, obtaining different SCF lexica in terms of size and accuracy.Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC\'12) url: http://www.lrec-conf.org/proceedings/lrec2012/summaries/390.html abstract: Lexica of predicate-argument structures constitute a useful tool for several tasks in NLP. This paper describes a web-service system for automatic acquisition of verb subcategorization frames (SCFs) from parsed data in Italian. The system acquires SCFs in an unsupervised manner. We created two gold standards for the evaluation of the system, the first by mixing together information from two lexica (one manually created and the second automatically acquired) and manual exploration of corpus data and the other annotating data extracted from a specialized corpus (environmental domain). Data filtering is accomplished by means of the maximum likelihood estimate (MLE). The evaluation phase has allowed us to identify the best empirical MLE threshold for the creation of a lexicon (P=0.653, R=0.557, F1=0.601). In addition to this, we assigned to the extracted entries of the lexicon a confidence score based on the relative frequency and evaluated the extractor on domain specific data. The confidence score will allow the final user to easily select the entries of the lexicon in terms of their reliability: one of the most interesting feature of this work is the possibility the final users have to customize the results of the SCF extractor, obtaining different SCF lexica in terms of size and accuracy. |

|  |
| --- |
| Nr 23 |
| Tipologia prodotto Contributo in Conferenza o Convegno (Senza ISBN) |
| Titolo How predictive are grammatical constructions in Italian? The case of the caused-motion construction |
| Descrizione |
| Elenco autori Quochi, Valeria |
| Ruolo svolto |
| anno pubblicazione 2012 |
| abstract: Differently from English, Italian has a rich morpho logical system and a relative free word-order. For these reasons, the suitability of a "full-scope" constructional approach to Italian is not given. Although Goldberg's (1995, 2006) version of Construction grammar language is constructions all the way down (or up), one could still argue that in Italian, i.e. a language rich in morphology, abstract, grammatical constructions do not play a role, thus weakening the constructionist view. One of the strong points in favour of goldberg's approach is that argument structure constructions in English have been found to be highly predictive of sentence meaning (Goldberg et al. 2005), which provides a motivationfor their early acquisition by children.  Many of such studies and evidences are still missing for Italian. This contribution will therefore attempt to start filling this gap by testing the predictive power of the Italian Caused Motion Construction. Data is taken from the CHILDES database (MacWhinney 2000) and annotated according to constructional properties and verb meaning. The annotation is then used to calculate the Cue and Category Validity (Murphy 2002) of both the Construction and the main verbs, which measures their predictive power (i.e respectively their reliability and availability) in relation to the overall sentence meaning. Results show that the Italian Caused Motion Construction is not only more reliable than verbs as a predictor of overall sentence meaning, but it is also more available. url: http://www.sle2012.eu/downloads/Book\_abstracts\_SLE2012\_23aug\_final.pdf abstract: Differently from English, Italian has a rich morpho logical system and a relative free word-order. For these reasons, the suitability of a \"full-scope\" constructional approach to Italian is not given. Although Goldberg\'s (1995, 2006) version of Construction grammar language is constructions all the way down (or up), one could still argue that in Italian, i.e. a language rich in morphology, abstract, grammatical constructions do not play a role, thus weakening the constructionist view. One of the strong points in favour of goldberg\'s approach is that argument structure constructions in English have been found to be highly predictive of sentence meaning (Goldberg et al. 2005), which provides a motivationfor their early acquisition by children.  Many of such studies and evidences are still missing for Italian. This contribution will therefore attempt to start filling this gap by testing the predictive power of the Italian Caused Motion Construction. Data is taken from the CHILDES database (MacWhinney 2000) and annotated according to constructional properties and verb meaning. The annotation is then used to calculate the Cue and Category Validity (Murphy 2002) of both the Construction and the main verbs, which measures their predictive power (i.e respectively their reliability and availability) in relation to the overall sentence meaning. Results show that the Italian Caused Motion Construction is not only more reliable than verbs as a predictor of overall sentence meaning, but it is also more available. |

|  |
| --- |
| Nr 24 |
| Tipologia prodotto Contributo in Conferenza o Convegno |
| Elenco autori Rubino, Francesco; Frontini, Francesca; Quochi, Valeria |
| Titolo Integrating NLP Tools in a Distributed Environment: A Case Study Chaining a Tagger with a Dependency Parser |
| Codice identificativo (ISBN) 9782951740877 |
| anno pubblicazione 2012 |
| abstract: The present paper tackles the issue of PoS tag conversion within the framework of a distributed web service platform for the automatic creation of language resources. PoS tagging is now considered a "solved problem"; yet, because of the differences in the tagsets, interchange of the various PoS taggers vailable is still hampered. In this paper we describe the implementation of a PoS-tagged-corpus converter, which is needed for chaining together in a workflow the FreeLing PoS tagger for Italian and the DESR dependency parser, given that these two tools have been developed independently. The conversion problems experienced during the implementation, related to the properties of the different tagsets and of tagset conversion in general, are discussed together with the solutions adopted. Finally, the converter is evaluated by assessing the impact of conversion on the performance of the dependency parser by comparing with the outcome of the native pipeline. From this we learn that in most cases parsing errors are due to actual tagging errors, and not to conversion itself. Besides, information on accuracy loss is an important feature in a distributed environment of (NLP) services, where users need to decide which services best suit their needsProceedings of the Eight International Conference on Language Resources and Evaluation (LREC\'12) url: http://www.lrec-conf.org/proceedings/lrec2012/summaries/726.html abstract: The present paper tackles the issue of PoS tag conversion within the framework of a distributed web service platform for the automatic creation of language resources. PoS tagging is now considered a \"solved problem\"; yet, because of the differences in the tagsets, interchange of the various PoS taggers vailable is still hampered. In this paper we describe the implementation of a PoS-tagged-corpus converter, which is needed for chaining together in a workflow the FreeLing PoS tagger for Italian and the DESR dependency parser, given that these two tools have been developed independently. The conversion problems experienced during the implementation, related to the properties of the different tagsets and of tagset conversion in general, are discussed together with the solutions adopted. Finally, the converter is evaluated by assessing the impact of conversion on the performance of the dependency parser by comparing with the outcome of the native pipeline. From this we learn that in most cases parsing errors are due to actual tagging errors, and not to conversion itself. Besides, information on accuracy loss is an important feature in a distributed environment of (NLP) services, where users need to decide which services best suit their needs |

|  |
| --- |
| Nr 25 |
| Tipologia prodotto Contributo in Conferenza o Convegno |
| Elenco autori Riccardo Del Gratta, Francesca Frontini, Monica Monachini, Valeria Quochi, Francesco Rubino, Matteo Abrate, Angelica Lo Duca |
| Titolo L-LEME: an Automatic Lexical Merger based on the LMF Standard |
| Codice identificativo (ISBN) 978-2-9517408-7-7 |
| anno pubblicazione 2012 |
| abstract: The present paper describes LMF LExical MErger (L-LEME), an architecture to combine two lexicons in order to obtain new resource(s). L-LEME relies on standards, thus exploiting the benefits of the ISO Lexical Markup Framework (LMF) to ensure interoperability. L-LEME is meant to be dynamic and heavily adaptable: it allows the users to configure it to meet their specific needs. The L-LEME architecture is composed of two main modules: the Mapper, which takes in input two lexicons A and B and a set of user-defined rules and instructions to guide the mapping process (Directives D) and gives in output all matching entries. The algorithm also calculates a cosine similarity score. The Builder takes in input the previous results, a set of Directives D1 and produces a new LMF lexicon C. The Directives allow the user to define its own building rules and different merging scenarios. L-LEME is applied to a specific concrete task within the PANACEA project, namely the merging of two Italian SubCategorization Frame (SCF) lexicons. The experiment is interesting in that A and B have different philosophies behind, being A built by human introspection and B automatically extracted. Ultimately, L-LEME has interesting repercussions in many language technology applicationsProceedings of the LREC 2012 Workshop on Language Resource Merging abstract: The present paper describes LMF LExical MErger (L-LEME), an architecture to combine two lexicons in order to obtain new resource(s). L-LEME relies on standards, thus exploiting the benefits of the ISO Lexical Markup Framework (LMF) to ensure interoperability. L-LEME is meant to be dynamic and heavily adaptable: it allows the users to configure it to meet their specific needs. The L-LEME architecture is composed of two main modules: the Mapper, which takes in input two lexicons A and B and a set of user-defined rules and instructions to guide the mapping process (Directives D) and gives in output all matching entries. The algorithm also calculates a cosine similarity score. The Builder takes in input the previous results, a set of Directives D1 and produces a new LMF lexicon C. The Directives allow the user to define its own building rules and different merging scenarios. L-LEME is applied to a specific concrete task within the PANACEA project, namely the merging of two Italian SubCategorization Frame (SCF) lexicons. The experiment is interesting in that A and B have different philosophies behind, being A built by human introspection and B automatically extracted. Ultimately, L-LEME has interesting repercussions in many language technology applications |

|  |
| --- |
| Nr 26 |
| Tipologia prodotto Contributo in Conferenza o Convegno |
| Elenco autori Poch, Marc; Toral, Antonio; Hamon, Olivier; Quochi, Valeria; Bel, Nuria |
| Titolo Towards a User-Friendly Platform for Building Language Resources based on Web Services |
| Codice identificativo (ISBN) 978-2-9517408-7-7 |
| anno pubblicazione 2012 |
| abstract: This paper presents the platform developed in the PANACEA project, a distributed factory that automates the stages involved in the acquisition, production, updating and maintenance of Language Resources required by Machine Translation and other Language Technologies. We adopt a set of tools that have been successfully used in the Bioinformatics field, they are adapted to the needs of our field and used to deploy web services, which can be combined to build more complex processing chains (workflows). This paper describes the platform and its different components (web services, registry, workflows, social network and interoperability). We demonstrate the scalability of the platform by carrying out a set of massive data experiments. Finally, a validation of the platform across a set of required criteria proves its usability for different types of users (non-technical users and providers).Proceedings of the Eighth International Conference on Language Resources and Evaluation, LREC 2012 url: http://www.lrec-conf.org/proceedings/lrec2012/pdf/543\_Paper.pdf abstract: This paper presents the platform developed in the PANACEA project, a distributed factory that automates the stages involved in the acquisition, production, updating and maintenance of Language Resources required by Machine Translation and other Language Technologies. We adopt a set of tools that have been successfully used in the Bioinformatics field, they are adapted to the needs of our field and used to deploy web services, which can be combined to build more complex processing chains (workflows). This paper describes the platform and its different components (web services, registry, workflows, social network and interoperability). We demonstrate the scalability of the platform by carrying out a set of massive data experiments. Finally, a validation of the platform across a set of required criteria proves its usability for different types of users (non-technical users and providers). |

|  |
| --- |
| Nr 27 |
| Tipologia prodotto Deliverable di Progetto |
| Titolo Final FLaReNet deliverable: Language Resources for the Future - The Future of Language Resources |
| Descrizione |
| Elenco autori Nicoletta Calzolari, Nuria Bel, Khalid Choukri, Joseph Mariani, Monica Monachini, Jan Odijk, Stelios Piperidis, Valeria Quochi, Claudia Soria |
| Ruolo svolto |
| anno pubblicazione 2011 |
| Altre informazioni |

|  |
| --- |
| Nr 28 |
| Tipologia prodotto Contributo in Conferenza o Convegno |
| Elenco autori Nicoletta Calzolari, Monica Monachini, Valeria Quochi |
| Titolo Interoperability Framework: The FLaReNet action plan proposal |
| Codice identificativo (ISBN) 978-974-466-564-5 |
| anno pubblicazione 2011 |
| abstract: Standards are fundamental to ex-change, preserve, maintain and integrate data and language resources, and as an essential basis of any language resource infrastructure. This paper promotes an Interoperability Framework as a dynamic environment of standards and guidelines, also intended to support the provision of language-(web)service interoperability. In the past two decades, the need to define common practices and formats for linguistic resources has been increasingly recognized and sought. Today open, collaborative, shared data is at the core of a sound language strategy, and standardisation is actively on the move. This paper first describes the current landscape of standards, and presents the major barriers to their adoption; then, it describes those scenarios that critically involve the use of standards and provide a strong motivation for their adoption; lastly, a series of actions and steps needed to operationalise standards and achieve a full interoperability for Language Resources and Technologies are proposed. abstract: Standards are fundamental to ex-change, preserve, maintain and integrate data and language resources, and as an essential basis of any language resource infrastructure. This paper promotes an Interoperability Framework as a dynamic environment of standards and guidelines, also intended to support the provision of language-(web)service interoperability. In the past two decades, the need to define common practices and formats for linguistic resources has been increasingly recognized and sought. Today open, collaborative, shared data is at the core of a sound language strategy, and standardisation is actively on the move. This paper first describes the current landscape of standards, and presents the major barriers to their adoption; then, it describes those scenarios that critically involve the use of standards and provide a strong motivation for their adoption; lastly, a series of actions and steps needed to operationalise standards and achieve a full interoperability for Language Resources and Technologies are proposed. |

|  |
| --- |
| Nr 30 |
| Tipologia prodotto Articolo in Rivista |
| Elenco autori Paul Thompson, John McNaught, Simonetta Montemagni, Nicoletta Calzolari, Riccardo del Gratta, Vivian Lee, Simone Marchi, Monica Monachini, Piotr Pezik, Valeria Quochi, CJ Rupp, Yutaka Sasaki, Giulia Venturi, Dietrich Rebholz-Schuhmann, Sophia Ananiadou |
| Titolo The BioLexicon: a large-scale terminological resource for biomedical text mining |
| Rivista BMC bioinformatics |
| Codice identificativo (ISSN) 1471-2105 |
| anno pubblicazione 2011 |
| Indice di classificazione |
| Impact Factor rivista |
| ruolo svolto |
| numero citazioni |
| Altre informazioni doi: 10.1186/1471-2105-12-397 url: http://www.biomedcentral.com/1471-2105/12/397 abstract: Background Due to the rapidly expanding body of biomedical literature, biologists require increasingly sophisticated and efficient systems to help them to search for relevant information. Such systems should account for the multiple written variants used to represent biomedical concepts, and allow the user to search for specific pieces of knowledge (or events) involving these concepts, e.g., protein-protein interactions. Such functionality requires access to detailed information about words used in the biomedical literature. Existing databases and ontologies often have a specific focus and are oriented towards human use. Consequently, biological knowledge is dispersed amongst many resources, which often do not attempt to account for the large and frequently changing set of variants that appear in the literature. Additionally, such resources typically do not provide information about how terms relate to each other in texts to describe events. Results This article provides an overview of the design, construction and evaluation of a large-scale lexical and conceptual resource for the biomedical domain, the BioLexicon. The resource can be exploited by text mining tools at several levels, e.g., part-of-speech tagging, recognition of biomedical entities, and the extraction of events in which they are involved. As such, the BioLexicon must account for real usage of words in biomedical texts. In particular, the BioLexicon gathers together different types of terms from several existing data resources into a single, unified repository, and augments them with new term variants automatically extracted from biomedical literature. Extraction of events is facilitated through the inclusion of biologically pertinent verbs (around which events are typically organized) together with information about typical patterns of grammatical and semantic behaviour, which are acquired from domain-specific texts. In order to foster interoperability, the BioLexicon is modelled using the Lexical Markup Framework, an ISO standard. Conclusions The BioLexicon contains over 2.2 M lexical entries and over 1.8 M terminological variants, as well as over 3.3 M semantic relations, including over 2 M synonymy relations. Its exploitation can benefit both application developers and users. We demonstrate some such benefits by describing integration of the resource into a number of different tools, and evaluating improvements in performance that this can bring. |

|  |
| --- |
| Nr 31 |
| Tipologia prodotto Contributo in Conferenza o Convegno (Senza ISBN) |
| Titolo The development of Light-'do' Verb Constructions in Italian |
| Descrizione |
| Elenco autori Quochi Valeria |
| Ruolo svolto |
| anno pubblicazione 2011 |
| abstract: This contribution presents the results of a study of the development of Light 'do' Verb Constructions in Italian based on naturalistic data. The claim is that there exists a Light Verb pivot schema that accounts for new productive formations and that this pattern is learnt by young children because it constitutes a labeling technique for naming new events, activities and situations. The findings of this research support two hypotheses of language acquisition. The results are based on analysis of longitudinal transcriptions of adult children interactions contained in the CHILDES databank (MacWhinney 2000).SLE 2011 - 44TH ANNUAL MEETING BOOK OF ABSTRACT url: http://sle2011.cilap.es/downloads/book\_abstracts.pdf abstract: This contribution presents the results of a study of the development of Light \'do\' Verb Constructions in Italian based on naturalistic data. The claim is that there exists a Light Verb pivot schema that accounts for new productive formations and that this pattern is learnt by young children because it constitutes a labeling technique for naming new events, activities and situations. The findings of this research support two hypotheses of language acquisition. The results are based on analysis of longitudinal transcriptions of adult children interactions contained in the CHILDES databank (MacWhinney 2000). |

|  |
| --- |
| Nr 32 |
| Tipologia prodotto Deliverable di Progetto |
| Titolo The FLaReNet Databook |
| Descrizione |
| Elenco autori Arranz V., Bel N., Budin G., Caselli T., Choukri K., Del Gratta R., Frontini F., Goggi S., Monachini M., Quochi V., Rubino F., Russo I. et alii |
| Ruolo svolto |
| anno pubblicazione 2011 |
| Altre informazioni |

|  |
| --- |
| Nr 33 |
| Tipologia prodotto Deliverable di Progetto |
| Titolo The Standards' Landscape Towards an Interoperability Framework |
| Descrizione |
| Elenco autori Monica Monachini, Valeria Quochi, Nicoletta Calzolari, Núria Bel, Gerhard Budin, Tommaso Caselli, Khalid Choukri, Gil Francopoulo, Erhard Hinrichs, Steven Krauwer, Lothar Lemnitzer, Joseph Mariani, Jan Odijk, Stelios Piperidis, Adam Przepiorkowski, Laurent Romary, Helmut Schmidt, Hans Uszkoreit, Peter Wittenburg |
| Ruolo svolto |
| anno pubblicazione 2011 |
| Altre informazioni |

|  |
| --- |
| Nr 34 |
| Tipologia prodotto Contributo in Conferenza o Convegno |
| Elenco autori Jezek E.; Quochi V. |
| Titolo Capturing Coercions in Texts: a First Annotation Exercise |
| Codice identificativo (ISBN) 2-9517408-6-7 |
| anno pubblicazione 2010 |
| abstract: In this paper we report the first results of an annotation exercise of argument coercion phenomena performed on Italian texts. Our corpus consists of ca 4000 sentences from the PAROLE sottoinsieme corpus (Bindi et al. 2000) annotated with Selection and Coercion relations among verb-noun pairs formatted in XML according to the Generative Lexicon Mark-up Language (GLML) format (Pustejovsky et al., 2008). For the purposes of coercion annotation, we selected 26 Italian verbs that impose semantic typing on their arguments in either Subject, Direct Object or Complement position. Every sentence of the corpus is annotated with the source type for the noun arguments by two annotators plus a judge. An overall agreement of 0.87 kappa indicates that the annotation methodology is reliable. A qualitative analysis of the results allows us to outline some suggestions for improvement of the task: 1) a different account of complex types for nouns has to be devised and 2) a more comprehensive account of coercion mechanisms requires annotation of the deeper meaning dimensions that are targeted in coercion operations, such as those captured by Qualia relations.Proceedings of the Seventh International Conference on Language Resources and Evaluation - LREC\'10 url: http://www.lrec-conf.org/proceedings/lrec2010/summaries/713.html abstract: In this paper we report the first results of an annotation exercise of argument coercion phenomena performed on Italian texts. Our corpus consists of ca 4000 sentences from the PAROLE sottoinsieme corpus (Bindi et al. 2000) annotated with Selection and Coercion relations among verb-noun pairs formatted in XML according to the Generative Lexicon Mark-up Language (GLML) format (Pustejovsky et al., 2008). For the purposes of coercion annotation, we selected 26 Italian verbs that impose semantic typing on their arguments in either Subject, Direct Object or Complement position. Every sentence of the corpus is annotated with the source type for the noun arguments by two annotators plus a judge. An overall agreement of 0.87 kappa indicates that the annotation methodology is reliable. A qualitative analysis of the results allows us to outline some suggestions for improvement of the task: 1) a different account of complex types for nouns has to be devised and 2) a more comprehensive account of coercion mechanisms requires annotation of the deeper meaning dimensions that are targeted in coercion operations, such as those captured by Qualia relations. |

|  |
| --- |
| Nr 35 |
| Tipologia prodotto Contributo in Conferenza o Convegno |
| Elenco autori Pustejovsky J.; Rumshisky A.; Plotnick A.; Jezek E.; Batiukova O.; Quochi V. |
| Titolo SemEval-2010 Task 7: Argument Selection and Coercion. |
| Codice identificativo (ISBN) 978-1-932432-70-1 |
| anno pubblicazione 2010 |
| abstract: The paper describes the Argument Selection and Coercion task for the SemEval-2010 evaluation exercise, which involves characterizing the type of compositional operation that exists between a predicate and the arguments it selects. Specifically, the goal is to identify whether the type that a verb selects is satisfied directly by the argument, or whether the argument must change type to satisfy the verb typing.Proceedings of the 5th International Workshop on Semantic Evaluation url: http://www.aclweb.org/anthology/S10-1005 abstract: The paper describes the Argument Selection and Coercion task for the SemEval-2010 evaluation exercise, which involves characterizing the type of compositional operation that exists between a predicate and the arguments it selects. Specifically, the goal is to identify whether the type that a verb selects is satisfied directly by the argument, or whether the argument must change type to satisfy the verb typing. |

|  |
| --- |
| Nr 36 |
| Tipologia prodotto Articolo in Rivista |
| Elenco autori Quochi V.; Del Gratta R.; Sassolini E.; Bartolini R.; Monachini M.; Calzolari N. |
| Titolo A Standard Lexical-Terminological Resource for the Bio Domain |
| Rivista Lecture notes in computer science |
| Codice identificativo (ISSN) 0302-9743 |
| anno pubblicazione 2009 |
| Indice di classificazione |
| Impact Factor rivista |
| ruolo svolto |
| numero citazioni |
| Altre informazioni doi: 10.1007/978-3-642-04235-5\_28 abstract: The present paper describes a large-scale lexical resource for the biology domain designed both for human and for machine use. This lexicon aims at semantic interoperability and extendability, through the adoption of ISO-LMF standard for lexical representation and through a granular and distributed encoding of relevant information. The first part of this contribution focuses on three aspects of the model that are of particular interest to the biology community: the treatment of term variants, the representation on bio events and the alignment with a domain ontology. The second part of the paper describes the physical implementation of the model: a relational database equipped with a set of automatic uploading procedures. Peculiarity of the BioLexicon is that it combines features of both terminologies and lexicons. A set verbs relevant for the domain is also represented with full details on their syntactic and semantic argument structure. |

|  |
| --- |
| Nr 38 |
| Tipologia prodotto Contributo in Conferenza o Convegno (Senza ISBN) |
| Titolo Relevance of Qualia Relations in Coercive Contexts |
| Descrizione |
| Elenco autori Jezek E.; Quochi V.; Calzolari N. |
| Ruolo svolto |
| anno pubblicazione 2009 |
| Altre informazioni |

|  |
| --- |
| Nr 39 |
| Tipologia prodotto Articolo in Rivista |
| Elenco autori Valeria Quochi |
| Titolo Usage scenarios and basic workflows |
| Rivista CLARIN Newsletter |
| Rivista CLARIN Newsletter |
| anno pubblicazione 2009 |
| Indice di classificazione |
| Impact Factor rivista |
| ruolo svolto |
| numero citazioni |
| Altre informazioni url: http://www.clarin.eu/sites/default/files/CLARIN\_Newsletter\_no\_6.pdf |

|  |
| --- |
| Nr 40 |
| Tipologia prodotto Contributo in Conferenza o Convegno |
| Elenco autori Quochi V.; Monachini M.; Del Gratta R.; Calzolari N. |
| Titolo A lexicon for biology and bioinformatics: the BOOTStrep experience |
| Codice identificativo (ISBN) 2-9517408-4-0 |
| anno pubblicazione 2008 |
| Altre informazioni url: http://www.lrec-conf.org/proceedings/lrec2008/pdf/576\_paper.pdf |

|  |
| --- |
| Nr 41 |
| Tipologia prodotto Contributo in Conferenza o Convegno |
| Elenco autori Quochi, Valeria; Calderone, Basilio |
| Titolo Learning properties of Noun Phrases: from data to functions |
| Codice identificativo (ISBN) 2-9517408-4-0 |
| anno pubblicazione 2008 |
| abstract: The paper presents two experiments of unsupervised classification of Italian noun phrases. The goal of the experiments is to identify the most prominent contextual properties that allow for a functional classification of noun phrases. For this purpose, we used a Self Organizing Map is trained with syntactically-annotated contexts containing noun phrases. The contexts are defined by means of a set of features representing morpho-syntactic properties of both nouns and their wider contexts. Two types of experiments have been run: one based on noun types and the other based on noun tokens. The results of the type simulation show that when frequency is the most prominent classification factor, the network isolates idiomatic or fixed phrases. The results of the token simulation experiment, instead, show that, of the 3 6 attributes represented in the original input matrix, only a few of them are prominent in the re-organization of the map. In particular, key features in the emergent macro-classification are the type of determiner and the grammatical number of the noun. An additional but not less interesting result is an organization into semantic/pragmatic micro-classes. In conclusions, our result confirm the relative prominence of determiner type and grammatical number in the task of noun (phrase) categorization. url: http://www.lrec-conf.org/proceedings/lrec2008/summaries/644.html abstract: The paper presents two experiments of unsupervised classification of Italian noun phrases. The goal of the experiments is to identify the most prominent contextual properties that allow for a functional classification of noun phrases. For this purpose, we used a Self Organizing Map is trained with syntactically-annotated contexts containing noun phrases. The contexts are defined by means of a set of features representing morpho-syntactic properties of both nouns and their wider contexts. Two types of experiments have been run: one based on noun types and the other based on noun tokens. The results of the type simulation show that when frequency is the most prominent classification factor, the network isolates idiomatic or fixed phrases. The results of the token simulation experiment, instead, show that, of the 3 6 attributes represented in the original input matrix, only a few of them are prominent in the re-organization of the map. In particular, key features in the emergent macro-classification are the type of determiner and the grammatical number of the noun. An additional but not less interesting result is an organization into semantic/pragmatic micro-classes. In conclusions, our result confirm the relative prominence of determiner type and grammatical number in the task of noun (phrase) categorization. |

|  |
| --- |
| Nr 42 |
| Tipologia prodotto Contributo in Conferenza o Convegno (Senza ISBN) |
| Titolo Using LMF to Shape a Lexicon for the Biomedical Domain |
| Descrizione |
| Elenco autori Monachini M.; Quochi V.; Del Gratta R.; Calzolari N. |
| Ruolo svolto |
| anno pubblicazione 2008 |
| abstract: This paper describes the design, implementation and population of the BioLexicon in the framework of BootStrep, an FP6 project. The BioLexicon (BL) is a lexical resource designed for text mining in the bio-domain. It has been conceived to meet both domain requirements and upcoming ISO standards for lexical representation. The data model and data categories are compliant to the ISO Lexical Markup Framework and the Data Category Registry. The BioLexicon integrates features of lexicons and terminologies: term entries (and variants) derived from existing resources are enriched with linguistic features, including sub-categorization and predicate-argument information, extracted from texts. Thus, it is an extendable resource. Furthermore, the lexical entries will be aligned to concepts in the BioOntology, the ontological resource of the project. The BL implementation is an extensible relational database with automatic population procedures. Population relies on a dedicated input data structure allowing to upload terms and their linguistic properties and "pull-and-push" them in the database. The BioLexicon teaches that the state-of-the-art is mature enough to aim at setting up a standard in this domain. Being conformant to lexical standards, the BioLexicon is interoperable and portable to other areas. abstract: This paper describes the design, implementation and population of the BioLexicon in the framework of BootStrep, an FP6 project. The BioLexicon (BL) is a lexical resource designed for text mining in the bio-domain. It has been conceived to meet both domain requirements and upcoming ISO standards for lexical representation. The data model and data categories are compliant to the ISO Lexical Markup Framework and the Data Category Registry. The BioLexicon integrates features of lexicons and terminologies: term entries (and variants) derived from existing resources are enriched with linguistic features, including sub-categorization and predicate-argument information, extracted from texts. Thus, it is an extendable resource. Furthermore, the lexical entries will be aligned to concepts in the BioOntology, the ontological resource of the project. The BL implementation is an extensible relational database with automatic population procedures. Population relies on a dedicated input data structure allowing to upload terms and their linguistic properties and \"pull-and-push\" them in the database. The BioLexicon teaches that the state-of-the-art is mature enough to aim at setting up a standard in this domain. Being conformant to lexical standards, the BioLexicon is interoperable and portable to other areas. |

|  |
| --- |
| Nr 43 |
| Tipologia prodotto Deliverable di Progetto |
| Titolo Bio-Lexicon DataBase: Architecture, Concepts and Loading Software |
| Descrizione |
| Elenco autori Del Gratta R., Monachini M., Quochi V., Sassolini E., Calzolari N. |
| Ruolo svolto |
| anno pubblicazione 2007 |
| Altre informazioni |

|  |
| --- |
| Nr 44 |
| Tipologia prodotto Contributo in Conferenza o Convegno (Senza ISBN) |
| Titolo BioLexicon Model and Implementation |
| Descrizione |
| Elenco autori Quochi V., Del Gratta R., Monachini M., Calzolari N. |
| Ruolo svolto |
| anno pubblicazione 2007 |
| Altre informazioni |

|  |
| --- |
| Nr 46 |
| Tipologia prodotto Contributo in Conferenza o Convegno (Senza ISBN) |
| Titolo Emergent Cognitive Functions of the Noun Phrase |
| Descrizione |
| Elenco autori Basilio Calderone; Valeria Quochi |
| Ruolo svolto |
| anno pubblicazione 2007 |
| Altre informazioni SLE 2007 Annual Meeting Book of Abstracts |

|  |
| --- |
| Nr 47 |
| Tipologia prodotto Deliverable di Progetto |
| Titolo FLaReNet: Fostering Language Resources Network. Grant Agreement n° 617001, eContentPlus |
| Descrizione |
| Elenco autori Calzolari N., Monachini M., Quochi V., Soria C., Goggi S., Baroni P. |
| Ruolo svolto |
| anno pubblicazione 2007 |
| Altre informazioni |

|  |
| --- |
| Nr 48 |
| Tipologia prodotto Contributo in Conferenza o Convegno (Senza ISBN) |
| Titolo Inferring the semantics of temporal prepositions in Italian |
| Descrizione |
| Elenco autori Caselli T., Quochi V. |
| Ruolo svolto |
| anno pubblicazione 2007 |
| Altre informazioni Proceedings of the Fourth ACL-SIGSEM Workshop on Prepositions url: http://www.aclweb.org/anthology/W07-1606 |

|  |
| --- |
| Nr 49 |
| Tipologia prodotto Contributo in Conferenza o Convegno (Senza ISBN) |
| Titolo Lexical Relations and Domain Knowledge: The BioLexicon Meets the Qualia Structure |
| Descrizione |
| Elenco autori Monachini M., Quochi V., Ruimy N., Calzolari N. |
| Ruolo svolto |
| anno pubblicazione 2007 |
| Altre informazioni |

|  |
| --- |
| Nr 50 |
| Tipologia prodotto Deliverable di Progetto |
| Titolo LocalBioLex: A database framework for biolinguistic research on integrated databases |
| Descrizione |
| Elenco autori Del Gratta R., Toral A., Quochi V., Monachini. M. |
| Ruolo svolto |
| anno pubblicazione 2007 |
| Altre informazioni |

|  |
| --- |
| Nr 51 |
| Tipologia prodotto Deliverable di Progetto |
| Titolo TimeML: An Ontological Mapping onto the UIMA Type Systems |
| Descrizione |
| Elenco autori Del Gratta R., Bartolini R., Caselli T., Enea A., Monachini M., Quochi V., Sassolini V. |
| Ruolo svolto |
| anno pubblicazione 2007 |
| Altre informazioni |

|  |
| --- |
| Nr 52 |
| Tipologia prodotto Contributo in Conferenza o Convegno |
| Elenco autori Quochi V.; Del Gratta R.; Sassolini E.; Monachini M.; Calzolari N. |
| Titolo Toward a Standard Lexical Resource in the Bio Domain |
| Codice identificativo (ISBN) 978-83-7177-413-3 |
| anno pubblicazione 2007 |
| abstract: The present paper describes a large-scale lexical resource for the biology domain designed both for human and for machine use. This lexicon aims at semantic interoperability and extendability, through the adoption of ISO-LMF standard for lexical representation and through a granular and distributed encoding of relevant information. The first part of this contribution focuses on three aspects of the model that are of particular interest to the biology community: the treatment of term variants, the representation on bio events and the alignment with a domain ontology. The second part of the paper describes the physical implementation of the model: a relational database equipped with a set of automatic uploading procedures. Peculiarity of the BioLexicon is that it combines features of both terminologies and lexicons. A set verbs relevant for the domain is also represented with full details on their syntactic and semantic argument structure. abstract: The present paper describes a large-scale lexical resource for the biology domain designed both for human and for machine use. This lexicon aims at semantic interoperability and extendability, through the adoption of ISO-LMF standard for lexical representation and through a granular and distributed encoding of relevant information. The first part of this contribution focuses on three aspects of the model that are of particular interest to the biology community: the treatment of term variants, the representation on bio events and the alignment with a domain ontology. The second part of the paper describes the physical implementation of the model: a relational database equipped with a set of automatic uploading procedures. Peculiarity of the BioLexicon is that it combines features of both terminologies and lexicons. A set verbs relevant for the domain is also represented with full details on their syntactic and semantic argument structure. |

|  |
| --- |
| Nr 53 |
| Tipologia prodotto Deliverable di Progetto |
| Titolo Bio-Lexicon Model and Preliminary ISO Conformant Data Categories |
| Descrizione |
| Elenco autori Quochi V., Monachini M., Calzolari N., Del Gratta R., Sassolini E. |
| Ruolo svolto |
| anno pubblicazione 2006 |
| Altre informazioni |

|  |
| --- |
| Nr 54 |
| Tipologia prodotto Contributo in Collana |
| Elenco autori Quochi, Valeria |
| Titolo Issues on the acquisition of Italian complex nominal from text corpora: a computational approach combining syntactic and semantic information |
| Codice identificativo (ISBN) 9788884922366 |
| anno pubblicazione 2005 |
| Altre informazioni |

|  |
| --- |
| Nr 55 |
| Tipologia prodotto Deliverable di Progetto |
| Titolo Final Resources Landscape |
| Descrizione |
| Elenco autori Baroni P., Calzolari N., Lenci A. , Quochi V., Ulivieri M. |
| Ruolo svolto |
| anno pubblicazione 2004 |
| Altre informazioni |

|  |
| --- |
| Nr 56 |
| Tipologia prodotto Contributo in Conferenza o Convegno |
| Elenco autori Quochi V. |
| Titolo Representing Italian Complex Nominals: A Pilot Study |
| Codice identificativo (ISBN) 2-9517408-1-6 |
| anno pubblicazione 2004 |
| abstract: A corpus-based investigation of Italian Complex Nominals (CNs), of the form N+PP, which aims at clarifying their syntactic and semantic constitution, is presented. The main goal is to find out useful parameters for their representation in a computational lexicon. As a reference model we have taken an implementation of Pustejovsky's Generative Lexicon Theory (1995), the SIMPLE Italian  Lexicon, and in particular the Extended Qualia Structure. Italian CN formation mainly exploits post-modification; of particular interest here are CNs of the kind N+PP since this syntactic pattern is highly productive in Italian and such CNs very often translate compound nouns of other languages. One of the major problems posed by CNs for interpretation is the retrieval or identification of the semantic  relation linking their components, which is (at least partially) implicit on the surface. Studying a small sample, we observed some interesting facts that could be useful when setting up a larger experiment to identify semantic relations and/or automatically learn the syntactic peculiarities of given semantic paradigms. Finally, a set of representational features exploiting the results from our corpus is  proposed.Proceedings of the Fourth International Conference on Language Resources and Evaluation, LREC\'04 abstract: A corpus-based investigation of Italian Complex Nominals (CNs), of the form N+PP, which aims at clarifying their syntactic and semantic constitution, is presented. The main goal is to find out useful parameters for their representation in a computational lexicon. As a reference model we have taken an implementation of Pustejovsky\'s Generative Lexicon Theory (1995), the SIMPLE Italian  Lexicon, and in particular the Extended Qualia Structure. Italian CN formation mainly exploits post-modification; of particular interest here are CNs of the kind N+PP since this syntactic pattern is highly productive in Italian and such CNs very often translate compound nouns of other languages. One of the major problems posed by CNs for interpretation is the retrieval or identification of the semantic  relation linking their components, which is (at least partially) implicit on the surface. Studying a small sample, we observed some interesting facts that could be useful when setting up a larger experiment to identify semantic relations and/or automatically learn the syntactic peculiarities of given semantic paradigms. Finally, a set of representational features exploiting the results from our corpus is  proposed. |

|  |
| --- |
| Nr 57 |
| Tipologia prodotto Deliverable di Progetto |
| Titolo \"Appendix F: Representing noun compounds and support verbs in MILE (PISA \& XMELLT)\" |
| Descrizione |
| Elenco autori Quochi, Valeria and Jan Odjik |
| Ruolo svolto |
| anno pubblicazione 2003 |
| Altre informazioni |

|  |
| --- |
| Nr 58 |
| Tipologia prodotto Contributo in Conferenza o Convegno (Senza ISBN) |
| Titolo Towards Multiword and Multilingual Lexicons: Between Theory and Practice |
| Descrizione |
| Elenco autori Calzolari N., Lenci A., Quochi V. |
| Ruolo svolto |
| anno pubblicazione 2003 |
| Altre informazioni |