Multilingual Event Detection using the NewsReader Pipelines

Rodrigo Agerri*, **Itziar Aldabe***, **Egoitz Laparra***, **German Rigau***
Antske Fokkens◇, **Paul Huijgen**◇, **Ruben Izquierdo**◇, **Marieke van Erp**◇, **Piek Vossen**◇
Anne-Lyse Minard∓, **Bernardo Magnini**∓
* University of the Basque Country (UPV/EHU), Spain
◇Vrije Universiteit Amsterdam, Netherlands
∓Fondazione Bruno Kessler, Italy

Abstract

We describe a novel modular system for cross-lingual event extraction for English, Spanish,, Dutch and Italian texts. The system consists of a ready-to-use modular set of advanced multilingual Natural Language Processing (NLP) tools. The pipeline integrates modules for basic NLP processing as well as more advanced tasks such as cross-lingual Named Entity Linking, Semantic Role Labeling and time normalization. Thus, our cross-lingual framework allows for the interoperable semantic interpretation of events, participants, locations and time, as well as the relations between them.

Keywords: Multilingual Event detection, Interoperable Semantic Processing, NLP Pipelines

1. Introduction

News texts report on events happening in the world. However, alternative sources may provide different perspectives on a specific topic. These differences can become particularly interesting when examining them across multiple sources and languages. For instance, they can contain redundant, incomplete or inconsistent information. Obviously, it is quite challenging to compare information from different sources, especially when they are written in different languages.

In this paper, we present a parallel architecture that largely apply the same linguistic analysis and produce the same language independent semantic representation. Our infrastructure currently integrates four complete NLP pipelines developed in the framework of the NewsReader project[1] for supporting event extraction in four different languages. The pipelines aim to identify *who* did *what*, *when* and *where* for texts written in English, Spanish, Dutch or Italian. The output of these individual pipelines is intended to be used as input for a system that obtains event centric knowledge graphs (Rospocher et al., 2016). As such, all pipelines have the same semantic core components for recognizing events, entities, concepts and time expressions, in order to extract the same language independent semantic representations.

Our pipelines are built as a *data centric architecture* so that modules can be adapted and replaced (even from alternative NLP toolkits and third party tools). All modules behave like Unix pipes: they all take standard input, do some annotation, and produce standard output which in turn is the input for the next module. Furthermore, its modular architecture allows for different configurations and for dynamic distribution of each module in independent machines boosting the performance of the whole when processing very large amounts of documents (Agerri et al., 2015).

2. Semantic interoperable framework

All modules included in the pipelines produce their output in the same format: the NLP Annotation Framework (NAF) (Fokkens et al., 2014). NAF is a standoff layered format for a host of different annotations, such as tokens, entities, predicates, semantic roles and time expressions.

Although NAF harmonizes the output of the different system modules, in order to achieve semantic interoperability, event information from multilingual sources, entity and event mentions are projected onto language independent knowledge representations. Thus, named entities are linked to English DBpedia entity identifiers through cross-lingual links existing to the Spanish, Italian and Dutch DBpedia counterparts. Nominal and verbal event mentions are aligned to abstract representations through the Predicate Matrix (López de Lacalle et al., 2014; López de Lacalle et al., 2016a; López de Lacalle et al., 2016b). Time expressions are all normalized to the ISO time format. Finally, we use the Collaborative Interlingual Index (CILI) to represent word senses (Vossen et al., 2016).

Consider the following English sentence from a Wikinews article:

> September 17, 2008
> Stock markets around the world, particularly those in the United States, have fallen dramatically today.

In this example, the expression *the United States* is detected as a named entity of the category LOCATION and is linked to the http://dbpedia.org/resource/United_States DBpedia entry. The predicate *fallen* and its corresponding argument *arg1* are linked to FrameNet (Baker et al., 1997), VerbNet (Kipper, 2005), PropBank (Palmer et al., 2005) and WordNet (Fellbaum, 1998) according to the predicate information included in the Predicate Matrix (López de Lacalle et al., 2014). The time expression *today* is normalized by reference to *2008-09-17* (the document creation time). Finally, *stock marked* is aligned to the concept ili-30-04323026-n.

Processing the Spanish, Dutch and Italian translation of the previous example through the corresponding pipelines results into the same language independent semantic representations.

[1] http://www.newsreader-project.eu

3. English pipeline

The English pipeline[2] currently provides the following linguistic annotations: document topic identification, Sentence segmentation, tokenization, Part of Speech (POS) tagging, Lemmatization, Named Entity Recognition and Classification (NERC), Constituent and Dependency Parsing, Nominal and Event Coreference Resolution, Word Sense Disambiguation (WSD), Named Entity Disambiguation (NED) and Wikification, Opinion mining, Semantic Role Labeling (SRL), extraction of Time expressions, Temporal Relations and Causal Relations, and Factuality detection.

IXA pipes[3] (Agerri et al., 2014) perform tokenization, POS tagging, lemmatization, NERC, constituent parsing and nominal coreference resolution.

DBpedia Spotlight (Mendes et al., 2011) is used to link the entities detected by the NERC module to DBpedia. Moreover, the pipeline also detects concepts that are relevant and they are not named entities using a wikification module. For example, given the example in Section 2., the pipeline detects *stock market* as a relevant concept appearing in DBpedia.

The SRL module detects predicates and roles of the sentences using the MATE-tools (Björkelund et al., 2010) and it also provides the corresponding interpretations in FrameNet, VerbNet, WordNet and ESO (Segers et al., 2015; Segers et al., 2016) using the Predicate Matrix (López de Lacalle et al., 2014).

Temporal processing aims at identifying temporal constraints of the events. It consists of time expression recognition and normalization, and temporal and causal relations extraction (Mirza and Minard, 2015; Mirza and Tonelli, 2014). In addition to the extraction of temporal relations as defined in TimeML, the module also identifies temporal anchoring of events, e.g. the date (explicit in the text or not) when an event took place or will occur.

We also identify whether an event is certain, probable or possible, whether it is confirmed or denied, or whether it takes place in the future or not. The core of the factuality module is trained on the factuality values from FactBank v1.0. A rule-based approach exploiting verbal morphology determines whether the event is situated in the future or not.

Document descriptors are useful in NewsReader to perform event coreference. The topic determines the domain of the document and this information, among other features, is used for event coreference resolution (Cybulska and Vossen, 2013). The module is based on the Multilingual Eurovoc thesaurus descriptors provided by the JRC Eurovoc Indexer JEX (Steinberger et al., 2012) also included in the pipeline.

4. Spanish pipeline

The NLP processing for Spanish[4] is similar to the English pipeline as they both share various modules to perform the processing: the JEX document topic identification module, the IXA-pipe modules and the WSD, NED, Wikification, SRL and Event coreference modules are created in the same manner. However, for Time expression detection and normalization we use HeidelTime (Strötgen et al., 2013), a multilingual temporal tagger. Identified temporal expressions are normalized and represented according to TIMEX annotations (Sundheim, 1996).

We are currently working on modules for factuality detection and temporal and causal relation extraction for Spanish. For the factuality module, we are using the SenSem corpus (Fernández-Montraveta and Vázquez, 2014).

5. Dutch pipeline

The Dutch pipeline[5] shares the IXA-pipe tokenizer and WSD tagger with the English and Spanish pipelines. As the Spanish pipeline, Heideltime is also used for detecting temporal expressions (van de Camp and Christiansen, 2013). For morpho-syntactic analysis, Alpino (van Noord et al., 2010) is used. The Dutch SRL module is a Python reimplementation of SoNar SRL (Clercq et al., 2012) for event predicates. As this SRL module does not handle nominalizations, we added a separate module to detect the predicates with part-of-speech noun and FrameNet Frames for one of their senses.

To enable the cross-lingual event mapping, links to common semantic resources are necessary. Since there are no predicate models for Dutch, we created a Dutch version of the PredicateMatrix by using the equivalence relations between the Dutch and English wordnets. If there is no match, we used the hypernym relations to infer Frames and Frame elements from hypernyms. We also exploited the cross-part-of-speech relation in the Dutch wordnet to obtain FrameNet data for deverbal nouns. In the Dutch pipeline, terms are enriched with synsets using the Dutch WSD module. For each synset, we integrate the PredicateMatrix data in the NAF output. Our SRL module outputs propBank roles for Dutch verbs. Since the Dutch Predicate Matrix provides mappings between Dutch and English predicates, PropBank and FrameNet roles, we select the most appropriate mappings for each predicate combining the scores of the WSD system for each synset, the Frames that are most dominant for each word and the Frame Elements that correspond with the PropBank roles in the SRL layer. These mappings are applied to the outcome of the SRL labeller and WSD system resulting in (typically) a set of matching FrameNet roles.

6. Italian pipeline

The Italian pipeline[6] is composed of modules from the TextPro tool suite (Pianta et al., 2008), extended by newly implemented modules and by third-party modules (DBpedia spotlight, also present in the English, Spanish, and Dutch pipelines).

As part of the NewsReader project we have developed modules for Time Processing in Italian (time expression extraction and normalization, event detection, temporal relation extraction, event factuality and predicate time anchor) (Mirza and Minard, 2014). They are based on the

[2] http://ixa2.si.ehu.es/nrdemo/demo.php
[3] http://ixa2.si.ehu.es/ixa-pipes/
[4] http://ixa2.si.ehu.es/nrdemo_es/demo.php
[5] http://kyoto.let.vu.nl/nwrdemo_nl/demo
[6] http://hlt-services2.fbk.eu:8080/nwrDemo/nwr

same methods as those used by the English modules, using language specific resources and training data.

Since no training annotated corpora exists for Italian SRL, we implemented a SRL system based on dependency relations (output of the dependency parser module), events (output of the event recognition module) and PropBank-like frames (built automatically using the MultiSemCor English-Italian aligned corpus (Bentivogli and Pianta, 2005)). In order to disambiguate predicate senses we use the version of the MultiWordNet (Pianta et al., 2002) provided by the Open Multilingual WordNet (Bond and Paik, 2012). Thus, predicates have external references to the Colaborative Interlingual Index (CILI). The match is created based on the lemma and morphological features, as well as comparing the roles extracted and those represented in the PropBank-like frames.

7. Evaluation

In order to assess the quality of the multilingual pipelines, the NewsReader project developed the MEANTIME corpus (Minard et al., 2016), a multilingual corpus containing intra-document and cross-document event annotations. The corpus is composed of 480 documents: 120 English wikinews articles around four topics: "Apple Inc.", "Airbus and Boeing", "General Motors", "Chrysler and Ford", and "Stock Market" and the translated versions of these articles into Dutch, Italian and Spanish. Translations have been done by professionals at sentence level. The creation of the corpus ensures access to freely available articles in all the languages and the option to compare the results of the NewsReader pipeline in the different languages at a fine-grained level.

We evaluated the English pipeline on standard datasets and in the MEANTIME corpus. Table 1 presents the results for NERC, nominal coreference, semantic role labeling, named entity disambiguation, temporal processing, factuality and event-coreference. On standard benchmark datasets our modules obtain state-of-the-art results. In the MEANTIME corpus the results are much lower. There are two main reasons for this: on the one hand, some of the modules have been trained on the same standard datasets. But more importantly, the standard corpora and the MEANTIME corpus differ on the annotation specification.

We also evaluated the Dutch pipeline on some standard datasets and in the MEANTIME corpus. Table 2 presents the results for NERC, semantic role labeling, named entity disambiguation and event-coreference. No results are provided for nominal coreference, temporal processing nor factuality. Again, our NERC module now on the Dutch standard benchmark dataset obtains very high results, improving the current state-of-the-art. As expected and for the same reasons explained before, in the MEANTIME corpus the results are much lower. However, we are now able to provide results for five different tasks. Compared to English, the Dutch results are lower. However, for NERC, the Dutch results are slightly better.

We also evaluated the Italian pipeline on some standard datasets and in the MEANTIME corpus. Table 3 presents the Italian results for NERC, semantic role labeling, named entity disambiguation, temporal processing, factuality and event-coreference. No results are provided for nominal coreference. As expected, in the MEANTIME corpus the results are much lower. However, we are now able to provide results for seven different tasks. Compared to English, the Italian results are lower for some tasks. However, for detecting time expressions, factuality and verbal coreference, the Italian results are slightly better.

We also evaluated the Spanish pipeline on some standard datasets and in the MEANTIME corpus. Table 4 presents the Spanish results for NERC, semantic role labeling, named entity disambiguation, temporal expressions and event-coreference. No results are provided for temporal relations and factuality. Again, as expected, in the MEANTIME corpus the results are much lower. However, we are now able to provide results for six different tasks. Compared to English, the Spanish results are lower. However, for NED, the Spanish results are slightly better.

In general, we observe very similar results across languages when having appropriate linguistic resources and annotation datasets. Detecting time relations and dealing with verbal coreference seems to be very difficult tasks. However, we present state-of-the-art results for all tasks. And for NERC, we improve the current state-of-the-art results.

In summary, we have presented a unique assessment exercise of the current NLP technology. As far as we know, we carried out the most complete and advanced multilingual evaluation of NLP pipelines.

8. Conclusions

In the NewsReader project we have developed four NLP pipelines for event extraction in English, Spanish, Dutch and Italian. The pipelines aim to identify *who* did *what*, *when* and *where* by adopting a common semantic representation. Semantic interoperability across the four languages is achieved by projecting entities, event predicates and roles, time expressions and concepts to language neutral semantic resources. We have evaluated the pipelines in standard datasets and in the MEANTIME multilingual corpus obtaining state-of-the-art results.

9. Acknowledgments

This work has been partially funded by NewsReader (FP7-ICT-2011-8-316404) and TUNER (TIN2015-65308-C5-1-R).

10. Bibliographical References

Agerri, R., Bermudez, J., and Rigau, G. (2014). IXA pipeline: Efficient and ready to use multilingual NLP tools. In *Proceedings of the 9th Language Resources and Evaluation Conference (LREC2014)*, Reykjavik, Iceland, May 26-31.

Agerri, R., Artola, X., Beloki, Z., Rigau, G., and Soroa, A. (2015). Big data for natural language processing: A streaming approach. *Knowledge-Based Systems*, 79(0):36 – 42.

Baker, C., Fillmore, C., and Lowe, J. (1997). The berkeley framenet project. In *COLING/ACL'98*, Montreal, Canada.

Bentivogli, L. and Pianta, E. (2005). Exploiting parallel texts in the creation of multilingual semantically anno-

Task	Evaluation metric	Standard Datasets		MEANTIME
		Dataset	F1	F1
NERC	CoNLL 2003	CoNLL 2003	91.18	68.67
Nominal coref	CoNLL 2011	CoNLL 2011	71.03	19.00
NED	F1	AIDA TAC 2011	77.66 68.92	68.58
SRL	CoNLL 2009	CoNLL 2009	84.74	34.74
Time exp.	TempEval3	Tempeval3	79.61	80.50
Temporal rel.	TempEval3	-	-	22.00
Factuality	Standard R	-	-	55.45 (R)
Event coref	F1	-	-	41.57

Table 1: English evaluation results on Standard benchmark datasets and NewsReader MEANTIME

Task	Evaluation metric	Standard Datasets		MEANTIME
		Dataset	F1	F1
NERC	CoNLL 2003	CONLL2002	85.04	70.24
Nominal coref	-	-	-	-
NED	Standard P & R	-	-	51.44
SRL	CoNLL 2009	-	-	26.76
Time exp.	TempEval3	-	-	58.70
Temporal rel.	-	-	-	-
Factuality	-	-	-	-
Event coref	F1	-	-	27.32

Table 2: Dutch evaluation results on Standard benchmark datasets and NewsReader MEANTIME

tated resources: The multisemcor corpus. *Nat. Lang. Eng.*, 11(3):247–261, September.

Björkelund, A., Bohnet, B., Hafdell, L., and Nugues, P. (2010). A high-performance syntactic and semantic dependency parser. In *Proceedings of the 23rd International Conference on Computational Linguistics: Demonstrations*, COLING '10, pages 33–36, Stroudsburg, PA, USA. Association for Computational Linguistics.

Bond, F. and Paik, K. (2012). A survey of wordnets and their licenses. In *Proceedings of the 6th Global WordNet Conference (GWC 2012)*, Matsue. 64–71.

Clercq, O. D., Hoste, V., and Monachesi, P. (2012). Evaluating automatic cross-domain semantic role annotation. In *Proceedings of the 8th International Conference on Language Resources and Evaluation Conference (LREC-2012)*, pages 88–93, Istanbul, Turkey.

Cybulska, A. and Vossen, P. (2013). Semantic relations between events and their time, locations and participants for event coreference resolution. In G. Angelova, et al., editors, *Proceedings of Recent Advances in Natural Language Processing (RANLP-2013)*, number ISSN 1313-8502, Hissar, Bulgaria, Sept 7-14. INCOMA Ltd.

C. Fellbaum, editor. (1998). *WordNet. An Electronic Lexical Database*. The MIT Press.

Fernández-Montraveta, A. and Vázquez, G. (2014). The sensem corpus: an annotated corpus for spanish and catalan with information about aspectuality, modality, polarity and factuality. *Corpus Linguistics and Linguistic Theory*, 10(2):273–288.

Fokkens, A., Soroa, A., Beloki, Z., Ockeloen, N., Rigau, G., van Hage, W. R., and Vossen, P. (2014). NAF and GAF: Linking linguistic annotations. In *Proceedings 10th Joint ISO-ACL SIGSEM Workshop on Interoperable Semantic Annotation*, page 9, Reykjavik, Iceland.

Kipper, K. (2005). *VerbNet: A broad-coverage, comprehensive verb lexicon*. Ph.D. thesis, University of Pennsylvania.

López de Lacalle, M., Laparra, E., and Rigau, G. (2014). Predicate matrix: Extending semlink through wordnet mappings. In *Proceedings of the 9th Language Resources and Evaluation Conference (LREC2014)*, Reykjavik, Iceland, May 26-31.

López de Lacalle, M., Laparra, E., Aldabe, I., and Rigau, G. (2016a). A multilingual predicate matrix. In *Proceedings of the 10th Language Resources and Evaluation Conference (LREC'16)*, Portorož, Slovenia.

López de Lacalle, M., Laparra, E., Aldabe, I., and Rigau, G. (2016b). Predicate matrix: automatically extending the semantic interoperability between predicate resources. *Language Resources and Evaluation*, pages 1–27.

Mendes, P. N., Jakob, M., García-Silva, A., and Bizer, C. (2011). Dbpedia spotlight: Shedding light on the web of documents. In *Proceedings of the 7th International Conference on Semantic Systems (I-SEMANTICS 2011)*, Graz, Austria, Sept. 7-9. ACM New York, NY, USA.

Minard, A.-L., Speranza, M., Urizar, R., na Altuna, B., van Erp, M., Schoen, A., and van Son, C. (2016). MEANTIME, the newsreader multilingual event and time corpus. In *Proceedings of the 10th Language Resources and Evaluation Conference (LREC'16)*, Portorož, Slovenia.

Mirza, P. and Minard, A.-L. (2014). FBK-HLT-time: a complete Italian Temporal Processing system for EVENTI-EVALITA 2014. In *Proceedings of the Fourth International Workshop EVALITA 2014*.

Mirza, P. and Minard, A.-L. (2015). Hlt-fbk: a complete

Task	Evaluation metric	Standard Datasets		MEANTIME
		Dataset	F1	F1
NERC	F1	Evalita 2007	82.10	56.77
Nominal coref	-	-	-	-
NED	Standard P & R	-	-	60.37
SRL	CoNLL 2009	-	-	31.62
Time exp.	TempEval3	Evalita 2014	82.7	85.7
Temporal rel.	F1	Evalita 2014	26.4	13.1
Factuality	Standard R	-	-	71.9
Event coref	F1	-	-	49.36

Table 3: Italian evaluation results on Standard benchmark datasets and NewsReader MEANTIME

Task	Evaluation metric	Standard Datasets		MEANTIME
		Dataset	F1	F1
NERC	CONLL2003	CONLL2002	84.16	65.54
Nominal coref	CONLL2011	SemEval 2010	64.22	15.74
NED	Standard P & R	TAC 2012	65.11	65.87
SRL	CoNLL 2009	CONLL2009	78.85	29.68
Time exp.	TempEval3	-	-	78.30
Temporal rel.	-	-	-	-
Factuality	-	-	-	-
Event coref	F1	-	-	30.37

Table 4: Spanish evaluation results on Standard benchmark datasets and NewsReader MEANTIME

temporal processing system for qa tempeval. In *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*, pages 801–805, Denver, Colorado, June. Association for Computational Linguistics.

Mirza, P. and Tonelli, S. (2014). An analysis of causality between events and its relation to temporal information. In *Proceedings of the 25th International Conference on Computational Linguistics (COLING2014)*, Dublin, Ireland, August 23-29.

Palmer, M., Gildea, D., and Kingsbury, P. (2005). The proposition bank: An annotated corpus of semantic roles. *Computational Linguistics*, 31(1):71–106, March.

Pianta, E., Bentivogli, L., and Girardi, C. (2002). Multiwordnet: developing an aligned multilingual database. In *Proceedings of the First International Conference on Global WordNet*, January.

Pianta, E., Girardi, C., and Zanoli, R. (2008). The textpro tool suite. In *Proceedings of LREC, 6th edition of the Language Resources and Evaluation Conference*, LREC-08, Stroudsburg, PA, USA. Association for Computational Linguistics.

Rospocher, M., van Erp, M., Vossen, P., Fokkens, A., Aldabe, I., Rigau, G., Soroa, A., Ploeger, T., and Bogaard, T. (2016). Building event-centric knowledge graphs from news. *Web Semantics: Science, Services and Agents on the World Wide Web*, In Press.

Segers, R., Vossen, P., Rospocher, M., Serafini, L., Laparra, E., and Rigau, G. (2015). ESO: a frame based ontology for events and implied situations. In *Proceedings of Maplex2015*.

Segers, R., Rospocher, M., Vossen, P., Laparra, E., Rigau, G., and Minard, A. (2016). The event and implied situation ontology (eso): Application and evaluation. In *Proceedings of the 10th Language Resources and Evaluation Conference (LREC'16)*, Portorož, Slovenia.

Steinberger, R., Ebrahim, M., and Turchi, M. (2012). Jrc eurovoc indexer jex - a freely available multi-label categorisation tool. In *LREC'12*, pages 798–805.

Strötgen, J., Zell, J., and Gertz, M. (2013). Heideltime: Tuning english and developing spanish resources for tempeval-3. In *Proceedings of the Seventh International Workshop on Semantic Evaluation, SemEval '13*, pages 15—19, Stroudsburg, PA, USA. Association for Computational Linguistics.

Sundheim, B. M. (1996). Overview of results of the muc-6 evaluation. In *Proceedings of a Workshop on Held at Vienna, Virginia: May 6-8, 1996*, TIPSTER '96, pages 423–442, Stroudsburg, PA, USA. Association for Computational Linguistics.

van de Camp, M. and Christiansen, H. (2013). Resolving relative time expressions in dutch text with constraint handling rules. In *Constraint Solving and Language Processing*, pages 166–177. Springer.

van Noord, G., Schuurman, I., and Bouma, G. (2010). Lassy syntactische annotatie. Technical report, Technical Report 19455, University of Groningen.

Vossen, P., Bond, F., and McCrae, J. P. (2016). Toward a truly multilingual global wordnet grid. In *Proceedings of workshop on the Collaborative Interlingual Index at the 8th Global WordNet Conference (GWC 2018)*, Bucharest, Rumania.