D5.2.1 Domain model for financial and economic events, version 1

Deliverable D5.2.1

Version FINAL

Authors: Marieke van Erp[1], Roxane Segers[1], Piek Vossen[1], Antske Fokkens[1], Filip Ilievski[1], Marco Rospocher[2], Rodrigo Agerri[3], Egoitz Laparra[3], German Rigau[3]

Affiliation: (1) VUA, (2) FBK, (3) EHU

Building structured event indexes of large volumes of financial and economic data for decision making

ICT 316404

Grant Agreement No.	316404
Project Acronym	NEWSREADER
Project Full Title	Building structured event indexes of large volumes of financial and economic data for decision making.
Funding Scheme	FP7-ICT-2011-8
Project Website	http://www.newsreader-project.eu/
Project Coordinator	Prof. dr. Piek T.J.M. Vossen VU University Amsterdam Tel. + 31 (0) 20 5986466 Fax. + 31 (0) 20 5986500 Email: piek.vossen@vu.nl
Document Number	Deliverable D5.2.1
Status & Version	FINAL
Contractual Date of Delivery	September 2014
Actual Date of Delivery	January 30, 2015
Type	Report
Security (distribution level)	Public
Number of Pages	77
WP Contributing to the Deliverable	WP05
WP Responsible	VUA
EC Project Officer	Susan Fraser

Authors: Marieke van Erp[1], Roxane Segers[1], Piek Vossen[1], Antske Fokkens[1], Filip Ilievski[1], Marco Rospocher[2], Rodrigo Agerri[3], Egoitz Laparra[3], German Rigau[3]

Keywords: Domain modelling, domain adaptation, event modelling, entity modelling

Abstract: This deliverable describes the first results on domain modelling for the financial-economic domain. We first demarcate the domain after which we provide in-depth analyses of the data sets that we have collected so far. These analyses aid us in assessing how much information about the domain is (implicitly) available within the domain and which information needs to be gathered from external sources. We then present our ontological domain models which provide the framework for our domain specific vocabularies. We also discuss how the domain model will be integrated with WP04 and WP05 and our future direction of work.

Table of Revisions

Version	Date	Description and reason	By	Affected sections
0.1	June 2014	Creation of document with structure	Marieke van Erp	ALL
0.2	August 2014	Added ESO description	Roxane Segers	4.3
0.3	September 2014	Added datasets analysis	Marieke van Erp	3
0.3	September 2014	Added ontological domain models	Marco Rospocher	5
0.4	September 2014	Restructured and integration domain model	Piek Vossen	7
0.5	September, 30, 2014	Various changes and completion of first complete draft after internal review	Piek Vossen, Roxane Segers, Marieke van Erp	ALL
0.6	11 January 2015	NED reranker	Marieke van Erp	7.1
0.7	25 January 2015	ESO based SRL post-processing	Filip Ilievski, Roxane Segers	6.1
0.8	25 January 2015	Predicate Matrix	Egoitz Laparra, German Rigau	6
0.9	25 January 2015	Internal review on new sections	VUA	6
0.9	30 January 2015	Check by coordinator	VUA	-

Executive Summary

This deliverable describes the first results on domain modelling for the financial-economic domain. We first demarcate the domain by providing a summary of our datasets which were previously described in Deliverables 1.1 "Definition of Data Sources" and 8.1 "Test Data and Scenarios". To assess how much information about the domain is already implicitly available from these datasets and how much and which information needs to be gathered from other resources, we provide some in-depth statistical analyses of our datasets. Our domain model is made up of two main parts, the ontological domain model and the vocabularies. The ontological domain model provides high level axioms that state the main concepts and the relationships between these concepts that exist in the domain. Our domain vocabularies provide the extension of these concepts by providing relevant instances of these concepts in our domain. By providing such vocabularies we can adapt the natural language processing tools in WP04 and WP05 to the domain, the details of which are presented in Section 7. We conclude with overall conclusions and our plans for future work.

Contents

Table of Revisions 3

1 Introduction 11

2 Scope of the Domain 13

3 Entities 13
- 3.1 Global Automotive Industry . . . 14
- 3.2 TechCrunch . . . 17

4 Events and roles 19
- 4.1 Designing the NewsReader Domain Ontology . . . 22
  - 4.1.1 How to represent an event instance and its corresponding situations 23
  - 4.1.2 Core classes and properties of the NewsReader Domain Ontology . 25
  - 4.1.3 Formalization of the rules for instantiating situations from events . 27
  - 4.1.4 Mappings from external resources to the NewsReader Domain Ontology 29
- 4.2 Building the Event and Situation Ontology version 1 . . . 29
  - 4.2.1 Hierarchy of events . . . 31
  - 4.2.2 Properties for defining pre and post conditions of an event . . . 32
  - 4.2.3 Roles for the entities affected by an event . . . 33
- 4.3 Event and Situation Ontology version 1 . . . 34

5 Vocabularies 41

6 Integrating ESO into the Predicate Matrix 43
- 6.1 SRL post-processing based on ESO . . . 43

7 Integration Domain Model 45
- 7.1 Domain adaptation of Natural Language Processing . . . 45
- 7.2 Domain adaptation of Event modeling . . . 49
- 7.3 Domain adapted interfacing . . . 53

8 Conclusions and Future Work 56

9 Appendix: automotive frames 60

10 Appendix: Examples of property value and property changes expressions 63

List of Tables

1 Types of mentions recognized in entity candidates occurring more than 150 times in Global Automotive Industry corpus . . . 16
2 Types of errors in entity mentions occurring more than 150 times in Global Automotive Industry corpus . . . 16
3 Types of mentions recognised in entity candidates occurring 50 or more times in the TechCrunch corpus . . . 18
4 Types of errors in entity mentions occurring 50 or more times in TechCrunch corpus . . . 18
5 Statistics on the predicates related to a FrameNet frame per predicate type 30
10 Mappings from ESO classes to SUMO classes and FrameNet frames . . . 37
6 Overview of static events . . . 39
7 Overview of situation rules for static and dynamic events . . . 39
8 Overview of Unary and Binary properties . . . 39
9 Overview of roles . . . 40
11 FrameNet Frame Entities reused for specifying the roles of entities . . . 40
12 Mapping between PropBank and ESO through FN. . . . 43
13 Mapping between PropBank and ESO through SUMO. . . . 43
14 Contextual frames occurring in the automotive industry data set . . . 61
15 Source introducing frames . . . 62
16 Grammatical event frames . . . 62

1 Introduction

The overall goal of the NewsReader project is to develop "deep-reading" technology for processing daily news streams across 4 languages. In the first year of the project, we developed generic NLP technology to extract the **who**, **what**, **when** and **where** from text. We applied the technology to 4 different data sets: 63K news articles on the automotive industry, 43K articles on the IT industry, 200K articles on the FIFA WorldCup and 19K articles from WikiNews on a variety of topics. We observed in the data that the generic technology makes considerable errors in terms of the correct interpretation of events and entities. We currently rely on the generic capacity of DBpedia Spotlight to detect the correct DBpedia URI and the Mate tools to select the correct predicate for the event and the corresponding roles. None of these tools were specifically developed for the data sets that we are dealing with. Finally, we attested the value of the rich event extraction for developing end user interfaces to the data. Although, events can be visualized in time lines and maps, this still does not provide sufficient insight in the implications of these changes. For example, the sentence "Apple hired Steve Jobs in 1995" implies that Steve Jobs did not work for Apple in some period before 1995 and did work in a period after 1995. To capture such implications, we need to have a more fine-grained model for a domain that defines the so-called pre and post conditions related to the events. Furthermore, it requires developing a type-hierarchy of events to group these conditions and map them to the proper vocabularies in languages. Event hierarchies and entity types in connection to properties and changes of values of properties also offers new ways of visualization of and interaction with the data.

This deliverable describes the creation of the first version of a domain model for the financial-economic domain. For this version of the model, we focus on learning which events and entities are important in the financial-economic domain (e.g. announcements, takeovers, product launches, market shares, companies, key persons etc.) from the data that was gathered and processed in NewsReader so far. The results of this inventory have been used to come to a domain ontology that captures the main concepts that play a role in the data. We also used the inventory to define a first set of important properties. We formalized the implications of the changes of the domain in terms of these properties so that we can support reasoning on the implications of the events. Furthermore, we differentiated between speech-act and cognition verb that are captured by the attribution model that we developed on the one hand and the so-called contextual events on the the other hand. Whereas the latter are domain-specific, the former occur in any domain and are interpreted through our forthcoming attribution model in RDF.[1]

In Section 3 we describe the types of entities and events that we can encounter and the different roles the entities can fulfill in relation to these events. This data is derived from the processed 63K documents in the automotive industry in relation to the DBpedia links that were found. Section 4 describes the types of events our current pipeline detects, which form the basis for the ontology. In Section 4.1, we describe the NewsReader ontology in

[1]This attribution model is further defined in the second part of year 2 of the project.

terms of its motivation and scope and in Sections 4.2 and 4.3 we describe the building of the ontology and the result. Section 5 provides the details on how the vocabularies are mapped to the ontology.

Section 7 explains how we exploit the domain adaptation in the event extraction and event modeling software developed in WP4, WP5 and WP7. In the case of the NLP modules, we show how the entity detection and the semantic role labeling are made sensitive to domain adaption. In the case of WP5, we show how reasoners can derive implications in relation to the detected events and add this to the RDF output in the KnowledgeStore. Finally, we describe the design of an interface that can exploit a data structure in which the implications for a domain are made explicit.

Next, we first define the boundaries of the domain and provide a summary of our datasets in Section 2.

2 Scope of the Domain

There are many aspects to the domain of financial and economic news. To optimise the coherence of our datasets, we chose two main subdomains, namely the Global Automotive Industry with sources from the LexisNexis database and the Technology Industry with open sources from the Web.

For the Global Automotive Industry domain, we gathered 6.1 million English documents from the LexisNexis database in which one or more car brands featured. The documents cover the period 2003-2013, which provides us with the lead-up to the 2007-2008 global financial economic crisis as well as its aftermath. This selection was further pared down to 64,540 articles for the first processing batch in the project's first year and only contains articles in which two or more car brands or persons strongly linked to them are mentioned.

For the Technology Industry domain, we collected 43,000 news articles from the TechCrunch website[2] as well as a data dump of the accompanying CrunchBase database[3] containing information about 180,000 companies and 200,000 biographies of people working in the Technology Industry domain. The TechCrunch articles were published between 2008 and 2013. The CrunchBase data dump was created at the same time as the TechCrunch articles were collected, and thus only contains updates until 2013.

Both the TechCrunch and Global Automotive Industry news articles were processed using the NewsReader NLP pipeline version 1, as described in Agerri *et al.* (2013). Through this pipeline, events, participants, locations and dates were recognized in the texts and linked to DBpedia entities. The first version of the NLP pipeline was set up as a baseline and was not geared specifically to the NewsReader domain. Through the analysis of entity recognition and (more specifically) linking (as presented in Section 3), we got some insights in the parts of our analyses that would benefit most from adapting the processing pipeline to the domain. Our analysis into the current event recognition and typing modules (Section 4) have resulted in a first domain ontology that is presented in Section 4.1.

3 Entities

In this section, we provide an analysis of the entities recognised and linked by the NLP modules in our datasets. In the analyses, we mostly focus on entity linking. Here, we aim to obtain insights into which entity mentions are incorrectly recognised, which links to external resources are lacking and – if there are links – whether they are correct or not.

With the completion of the benchmark dataset over the summer of 2014, a quantitative analysis of the NLP modules in the pipeline is scheduled for the autumn of 2014. This means that we cannot yet get a full overview of the precision and recall of the named entity module, but a qualitative analysis of the obtained named entities and their links already provides some interesting observations.

[2] `http://www.techcrunch.com`
[3] `http://www.crunchbase.com`

The NewsReader NER module uses machine learning and is trained on CoNLL 2003 data (Tjong Kim Sang and De Meulder, 2003), which was annotated according to the MUC guidelines (Chinchor and Robinson, 1997). In the first version of the NWR pipeline, 4 entity classes were discerned: Person, Location, Organisation and Miscellaneous. These classes are quite coarse-grained and, as the analyses of our datasets shows, sometimes too coarse-grained.

For the analyses of our datasets, we have chosen to focus on the most commonly occurring entity instance according to our pipeline. For the global automotive industry, we looked at the top 500 occurring entities for which a DBpedia link was found, and the top 147 instances (occurring more than 150 times) for which no link was found. For the TechCrunch data, the top 200 entity instances with DBpedia links and the top 149 instances without DBpedia links were analysed. We have divided our analysis of the mentions with links into those that are out of domain, error in the named entity recognition module, and other. For those cases in which the modules did not find a link, we classified the errors in the following categories: entity not present in DBpedia, spelling variation, not an entity or an event, conjunction, and other. In the remainder of this section, we describe our analyses as well as recommendations on how to overcome the domain specific errors.

3.1 Global Automotive Industry

In the 64,540 Global Automotive Industry articles, the NLP pipeline detected 2,538,046 entity mentions, which were aggregated into 344,251 unique entities. For 143,050 of these unique entities, the named entity linking module was able to find a link to DBpedia resources, covering 2,023,826 (79.74%) mentions.

We inspected the links for the 500 most commonly occurring entity mentions and found that the majority (446/1,571,831 mentions) were correct. The linking module also manages to correctly link 'Nissan Diesel' to `http://dbpedia.org/resource/UD_Trucks` (formerly known as Nissan Diesel) and 'LRX' to `http://dbpedia.org/resource/Range_Rover_Evoque` (LRX was the name of the prototype, it was brought to the market as the Evoque model).

The 54 (88,670 mentions) incorrectly linked entity instances can be classified into 3 different categories:

Out of domain link (20 instances) If an entity outside the car domain is more prevalent in DBpedia, that may get chosen over the car domain entity that is meant in the text or if the entity that is meant is not present in DBpedia but it is similar enough to an entity in the car domain to get linked. E.g. 'Lincoln' is linked to `http://dbpedia.org/resource/Abraham_Lincoln` instead of `http://dbpedia.org/page/Lincoln_Motor_Company`.

Error in Named Entity Recognition (29 instances) Mostly incorrect entity boundaries are recognised, such as 'D.', which is linked to `http://dbpedia.org/resource/Democratic_Party_(United_States)` or otherwise non-entities such as 'internet'.

Other (5 instances) The mention 'Ford' is linked to different resources such as `http://dbpedia.org/resource/Ford_Motor_Credit_Company`, `http://dbpedia.org/resource/Henry_Ford` and `http://dbpedia.org/resource/William_Clay_Ford,_Jr.` and it is unclear why exactly a particular resource is chosen over another in each instance. All resources are relevant to the domain.

We are investigating adapting the linking module to utilise more background knowledge from the car domain by incorporating knowledge about commonly co-occurring entities in the domain. However, a more pressing problem are the entities for which the module could not find a link. In our corpus, there were 514,220 mentions (201,201 unique mentions) for which the named entity linking module could not find a link to DBpedia. We inspected the mentions that occurred more than 150 times in our dataset (covering 147 unique entity instances or 45,463 mentions).

We first look at the types of entities that are recognised in the dataset but could not be linked to a DBpedia entry. As can be seen in Table 1, most of the entities in the global automotive industry are of type organisation or person, but products (such as car models) are also mentioned in a number of cases. This is currently not a type of entity that is recognised by the NLP pipeline, although it does provide very valuable information about the domain; we therefore aim to include it in the next version of the pipeline.

Of the 35 instances that the pipeline classified as entities, 17 cases are parts of entities (e.g., "Fuji Heavy" instead of "Fuji Heavy Industries" or "PSA Peugeot" instead of "PSA Peugeot Citroën"). This could be resolved in the processing pipeline by a feedback loop to the chunker which ought to be able to detect the correct chunk boundaries.

A similar number of instances (18) concerns cases that should not have been flagged as entities or parts of entities such as "upstairs loft", "Copyright AFX News Limited" or "Wheel drive". Filtering these out is a bit more difficult, but a post-processing rule for entity mentions could require that the content words in the entity mention are capitalised. Further, terms such as 'Copyright' ought not to be included.

There is also one instance of a conjunction that contains two types of entities, that is "Porsche and Piech", which should have been recognised as two separate entities, this is a difficult issue as there are also numerous entities that are made up of conjunctions ("Barnes and Noble", "Standard & Poor" etc.). However, domain information can teach our system that "Porsche" as well as "Piech" on their own are a very important entities in the car domain, enabling us to remedy some of these cases. "Porsche and Piech" is not the only conjunction that comes up in that analysed data, but it is the only one in which two different entity types are present. As Table 2 shows, there are 4 more unique conjunctions to be found, but these all contain two entities of the same type such as "Jaguar and Land Rover", which are both organisations.

When we look at the entities that are correctly recognised by the Named Entity Recognition module, but could not be linked to DBpedia, a major part of the problem is caused by the entity not being present in DBpedia (36.12%). In a majority of the cases (26), it concerns an unknown person, for example an industry analyst who is mentioned in a quote in an article. There are also 18 small companies mentioned such as local funeral homes

Type	# Mentions	# Unique
Person	12,084 (26.58%)	43
Organisation	16,217 (35.67%)	61
Person/Organisation	233 (0.51%)	1
Location	874(1.9%)	3
Product	1,536 (3.38%)	6
Not an entity	14,519(31.94%)	35

Table 1: Types of mentions recognized in entity candidates occurring more than 150 times in Global Automotive Industry corpus

Type	# Mentions	# Unique
Entity not present in DBpedia	16,423 (36.12%)	64
Spelling variation/nickname	11,417 (25.11%)	38
Not an entity or event	14,519 (31.94%)	35
Conjunction	1,716(3.77%)	5
Recall error	1,388 (3.05%)	5

Table 2: Types of errors in entity mentions occurring more than 150 times in Global Automotive Industry corpus

or industry magazines. There are also a few organisations mentioned that are not present in the english DBpedia, which was consulted for the English language domain, but that are present in the Dutch, French and or German version of DBpedia (such as "APEAL", the Association of European Producers of Steel for Packaging). Linking all English entity mentions to DBpedia version in different languages is not desirable since this is not a very prevalent issue. Rather, it makes more sense to try to include a database that contains more information about organisations and persons specific in the cars domain. Some initial analyses of the OpenCorporates.com[4] website lead us to believe that such a resource might enable us to link a large part of the organisation mentions not found in DBpedia to an external resource. However, the problem for the persons still exists, as it is unlikely that we can find one or more resources that cover all persons mentioned. In that case, we aim to create instances within the NewsReader data space for these entities, similar to how identifiers for event instances are created.

We also encounter a fairly substantial set of entities (14 unique) that concern subdivisions of the bigger companies (e.g., "Ford Sollers", "Volkswagen Canada", "Benz Cars"). These types of metonymic relations are difficult to recognise, but as with the conjunctions (e.g. "Jaguar and Land Rover"), we might be able to split the mention and provide a link to the head of the mention, which in many cases would be the organisation, possibly this link could be typed as a "part of" relation, but for this the data model would have to be adjusted. This would only be a viable option if the linking module can identify a relevant link with a high confidence, in order to not introduce too many erroneous links.

[4] `https://opencorporates.com/`

38 unique instances could not be matched to DBpedia because of a spelling variation or nickname, for example a person was referred to only by his/her last name instead of his/her full name ("Zetsche" instead of "Dieter Zetsche"). When we look beyond the entities that occur more than 150 times in the dataset, we find even more variations including titles ("Dr. Dieter Zetsche") or full job titles ("DaimlerChrysler Chairman Dieter Zetsche"). Rules for splitting titles and recognising job titles could be used to mitigate this issue. Furthermore, we can also build a resource of last names that are common in our dataset and perform a coreference step on these to link the different entity mention variations to a single entity instance.

The last error type we find is a recall error, in which we know the entity exists in DBpedia, except no link is made. This could be caused by the entity linking module not having a high enough confidence to link the entity or an error in the data conversion ("Chrysler Town & Country"). The former is a matter of finetuning the confidence in the module, the latter of improving the initial data cleanup (from XML to raw text).

3.2 TechCrunch

In the 43,000 TechCrunch articles, the NLP pipeline detected 807,088 entity mentions, which were aggregated into 212,611 unique entities. For 102,141 entities, links to DBpedia resources were identified, covering 608,801 (75.43%) of the entity mentions.

We inspected the links for the 200 (222,467 mentions) most commonly occurring entity instances, and found that for 185 of the entities (212,133 mentions) the correct link had been identified. We categorised the 15 cases (10,334 mentions) in which an incorrect link was provided as following:

Out of domain link (10 instances) If an entity outside the technology domain is more prevalent in DBpedia, that may get chosen over the technology domain entity that is meant in the text or if the entity that is meant is not present in DBpedia but it is similar enough to an entity in the car domain to get linked. E.g. 'Box' is linked to `http://dbpedia.org/resource/Box` instead of `http://dbpedia.org/page/Box_(company)`.

Error in Named Entity Recognition (5 instances) Mostly incorrect entity boundaries are recognised, such as 'no.', which is linked to `http://dbpedia.org/resource/Norwegian_language`).

We aim to take the same approach as in the global automotive industry domain here, by adapting the linking module to give preference over in-domain entities.

There were 198,287 mentions (110,470 unique) for which the named entity linking module could not find a link to DBpedia. We inspected the mentions that occurred more than 50 times in our dataset (covering 149 unique entities or 9,986 mentions).

As can be seen from Tables 3 and 1, the TechCrunch dataset has quite a different distribution of entity types from the Global Automotive Industry dataset. For the TechCrunch

Type	# Mentions	# Unique
Person	2,156 (21.59%)	25
Organisation	4,222 (42.28%)	44
Location	84 (0.84%)	1
Product	2,025 (20.28%)	21
Event	775 (7.76%)	8
Not an entity or event	724 (7.2%)	10

Table 3: Types of mentions recognised in entity candidates occurring 50 or more times in the TechCrunch corpus

Type	# Mentions	# Unique
Entity not present in DBpedia	7,776 (77.86%)	85
Spelling variation/nickname	576 (5.76%)	4
Recall error	135 (1.35%)	2
Event	775 (7.76%)	8
Not an entity or event	724 (7.25%)	10

Table 4: Types of errors in entity mentions occurring 50 or more times in TechCrunch corpus

dataset, we also chose to include events that were labelled as entities in our analysis as they occurred in almost 8% of the analysed cases. To start with the events, these are often nominal events, a class of events that is currently not recognised in the NLP pipeline. These events, such as "Startup Alley" and "Demo Day" are very specific to the domain and behave very much like named entities and it is thus not surprising that they are labelled as such. From a domain specific resource such as CrunchBase,[5] we can infer for some of these that they are not entities as they are present but they are not labelled as persons or organisations.

Another important category in the TechCrunch domain are products, which make up a far larger part of the entities mentioned than in the Global Automotive Industry corpus (20.28% vs 3.38%). The reason for this is that the main topic of the news articles in the technology domain is concerned with new products, whereas product releases are far less frequent in the automotive domain. There the majority of the news articles concerns interactions between companies and people involved in those companies.

As the technology domain is made up of small startups that either make it or break, they and their products are less established and thus less likely to occur in DBpedia, which is the case in 77.86% or 85 unique entities of those analysed. We found that a significant part of the entities could be linked to the CrunchBase resource (54 unique entities, 62.56% of the mentions that were not found in DBpedia). This resource also provides us with biographical information about persons and company histories, therefore linking entities to CrunchBase seems a viable option for this domain.

[5] `http://www.crunchbase.com`

There are also spelling variations and nicknames to be found in this domain. The types of variations are of a different type than in the Global Automotive Industry, namely they seem more informal. We find for example “Mike Arrington” instead of “Michael Arrington” and “Zuck” as shorthand for “Mark Zuckerberg”. Nicknames will be difficult to resolve, but spelling variations could be resolved with a list of common names that specifies that “Bill” is often used as an informal form for “William” etc.

Our analyses indicate that a majority of the problems for the technology domain lie in the lack of coverage by DBpedia. We will therefore look into adding other resources containing information about entities in the domain to our NLP pipeline in order to improve our linking recall. Below we summarise our recommendations for adapting our entity recognition and linking modules.

Recommendations for Domain Adaptation Entities

Based on the analyses presented in this section, the following steps can be undertaken to better recognise and link entities in the domain:

Give preference to in-domain entities to overcome entities being linked to the most popular entities in the general domain;

Link to additional resources to overcome gaps in coverage of DBpedia;

Include product category in particular for the technology domain this makes up a major part of the relevant entities;

In-domain coreference resolution to be able to link different variants of the same name;

Domain specific post-processing rules to split up conjunctions and deal with chunking. errors

In Section 7 we further describe how these recommendations are incorporated into the NLP pipeline.

4 Events and roles

Just as for entities, the fact that we use generic modules and resources for a specific domain is expected to have an impact on the quality of the interpretation of events and the roles that entities play in these events. These errors relate to:

- Detecting mentions of events: i.e. words and expressions are missed as referring to events or wrongly assigned to events;
- Determining the meaning of words and expressions: i.e. the wrong type of event is assigned or no sense is assigned so that the type cannot be derived;

- Detecting the words and expressions that express roles related to an event;
- Determining the type of role that some participant or concept plays in relation to the event.

However as long as there has not been any benchmarking against annotated domain data, we will not know the exact performance of the generic modules on the domain data. Such quantitative benchmarking is planned in the second part of year 2 of the project after completion of the annotation. Nevertheless, just as for entities, we can carry out a statistical analysis of the processed data to obtain overall insights in the type of errors that are made and to get an idea of the salient domain concepts (events and roles). Such an analysis can form the basis for a domain model to add precision and recall to the modules and to exploit the data in other ways.

Typically, the range of events (e.g. meanings of predicates), the possible roles that can occur with these events and the structural constraints are defined in the lexical resources and annotated corpora that are used for training the modules. In our case, the PredicateMatrix (de Lacalle *et al.*, 2014) is used as the basic resource to define the types of events and their roles. Word meanings, event types and roles are defined through links with VerbNet (Kipper *et al.*, 2000), PropBank (Palmer *et al.*, 2005), NomBank (Meyers *et al.*, 2004), WordNet (Fellbaum, 1998) and FrameNet (Baker *et al.*, 1998). The Mate tools (Björkelund *et al.*, 2009) are used to assign predicates and roles to expressions in context with a specification of the PredicateMatrix types that match the interpretation. Adaptation to the domain can potentially be done by improving the PredicateMatrix on the one hand (e.g. by extending the coverage to domain specific words and expressions) and providing domain specific training data for Mate tools on the other hand.

There are however a number of other reasons why a more detailed and specific adaptation to the domain is required. The events types and the roles coming from resources such as PropBank, VerbNet and FrameNet are not very well structured. There is a plethora of classes with unclear relations to each other. A resource such as FrameNet resembles more a graph than a well structured thesaurus or ontology. If not for the event types and roles derived from the PredicateMatrix, the data can only be accessed through the words used in the textual sources.

At the hackathon in London in June 2014, it became apparent that there is a great need for a more adequate hierarchical organization of events in classes or types. The general feeling observed is that the variation of words and expressions to express similar information and the ambiguity of these expression severely hamper the querying and interfacing the data. This will become even more complex if text in other languages than English are processed to provide content. A language neutral classification of the events and their roles helps querying the data and provide a better statistical insights in the events that took place and what trends may be shown. Similarly, the design of the Decision Support Tool Suite (van Hage and Ploeger, 2013) requires some hierarchical structuring of events and participants to allow for generalizations and for effectively interacting with large data structures, i.e. querying and linking over categories of events and entities rather than individual lexicalizations and wordings.

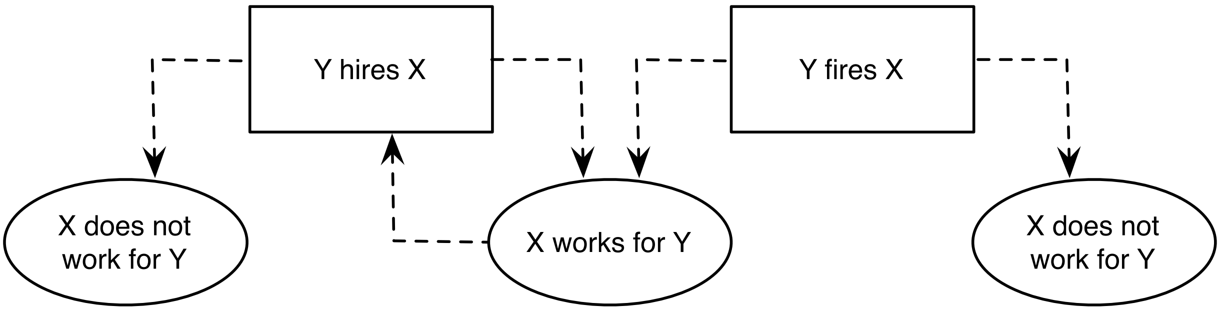

Figure 1: Sample events and situation-chain in the NewsReader domain

Currently, most of our data are presented as events with participants. In the case of the 63K car documents (1% of the total), we extracted about 1.7 millions events involving hundreds of thousands of participants. These events can be sorted in time and or by participant but that does not explain what are the implications of these events and how we can effectively present event sequences to users. In the current version of the NWR pipeline, direct statements for changes (e.g. X is fired) and situations (X works for Y) are made, but models for linking these implications are missing. Several technical and user-meetings carried out during the second year of the project made clear that our representation of events would benefit from a model that makes these implications explicit. We therefore concluded that we need a domain model that explicitly defines events in terms of the implied changes in the world: every event has a formally defined pre-situation and post-situation. Situations are entities for which hold that some set of statements is true for some period of time. Such an event and situation model is illustrated by Figure 1. This figure represents a chain of events (boxes) and situations (ovals). Each event can be associated with two situations that will hold before and after the event. In the case that 'Y fires X', we want to infer that 'X was working for Y' before this event and 'X is not working for Y' after the event.[6]

By explicitly modeling the situations pertaining to events, we can extract sequences of situations and changes over time regardless of if this information was directly expressed in text, or inferred by some reasoner on the basis of the events extracted from the text. Such a reasoner needs an ontology to link dynamic events to implied changes in values of properties.

The above observations led us to the decision to build a NewsReader Event and Situation Ontology for the financial-economic domain. This Event and Situation Ontology (ESO) is thus based on the following requirements:

- enabling to add more precision and recall to the detection of events and roles in the domain;

[6]Implications from situations to events are more complicated; in this case we can only state that if X works for Y, then in some point of time he must have been hired.

- to differentiate events in terms of relevance for the domain;
- to allow for a more proper interfacing to the data graphs through hierarchical organization;
- to allow for more efficient querying the data by grouping expressions of similar events to their corresponding types;
- to allow for reasoning over implication of changes and interfacing more properly to these changes from a user perspective.

In addition to these user requirements, the domain ontology should also be compliant to the following requirements:

- follow best practices and standards of formal ontology design;
- re-use as many existing resources as possible (lexicons and ontologies);
- enable transfer of the ontology to other languages than English.

In the next section 4.1, we will first describe the ontology meta model that we designed. We then provide a statistical and qualitative analysis of the events and roles that occurred in the car data set in section 3. This provided the basis for building the first version of the domain ontology, the process of which is described in section 4.2. The resulting contents of the current version are described in more detail in section 4.3

4.1 Designing the NewsReader Domain Ontology

In NewsReader, we are interested in representing events and their effects on entities involved in them. For instance, a "firing" event, where a company fires one of its employee, marks a change in an employment condition: before the event (aka, *pre-situation*), the employee is employed at the company, while after the event (aka, *post-situation*) the employee is no more employed at the company.

To be able to represent events and situations, the NewsReader Domain Ontology will define two main classes of entities: events and situations. An *event* is an entity that describes some change in the world. It has participants and a time (interval) associated to it. An event exists independently from the fact that it actually happens (e.g., hypothetical events). Typically, an event is associated with two situations: the situation before the event (pre-situation) and the one after the event (post-situation). The effects of an event are described in terms of the statements that hold in the situations associated to the event.

If we consider for instance a firing event:

In 2012, employeeA and employeeB were fired by companyA

we can identify a pre-situation (i.e., before the event):

employeeA works for companyA
employeeB works for companyA

as well as a post-situation (i.e., after the event):

employeeA does not work for companyA
employeeB does not work for companyA

A *situation* is an entity which is associated with a period of time where a set of statements (aka *fluents* in situation calculus) are true. It is a partial and "perspectival" description of the state of the world during the period of time it is associated with. It is partial because it does not describe the totality of propositions that are true in the world during the period of time associated to the situation. It is perspectival because it describes the point of view of a particular "agent".

4.1.1 How to represent an event instance and its corresponding situations

In the original situation calculus the predicate "holdsAt($r(a, b), s$)" is used to model the fact that "a and b are related with the relation r in situation s". In our proposal, we adopt recent advances in Semantic Web technologies, relying on the notion of "named graph": a named graph will be associated to each situation s, and it will contain all triples a, R, b holding in it.

Let's consider the aforementioned firing event example. The SRL module of the NewsReader pipeline will annotate the sentence "In 2012, employeeA and employeeB were fired by companyA" with the following information:

- fired → frame fn:firing;
- employeeA → frame element fn:Employee of frame fn:Firing;
- employeeB → frame element fn:Employee of frame fn:Firing;
- companyA → frame element fn:Employer of frame fn:Firing;

In addition, a time expression will be associated to the term "in 2012".

From this linguistic annotations, we will instantiate some individuals and assertions on them to formally represent the event according to standard Semantic Web formalisms. In details, we will instantiate a named graph of the form

```
:obj-graph-eventX {
    :eventX
        a                                    nwr:LeavingAnOrganization ;
        nwr:LeavingAnOrganization_employee   :employeeA                ;
        nwr:LeavingAnOrganization_employee   :employeeB                ;
        nwr:LeavingAnOrganization_employer   :companyA                 ;
        sem:hasTime                          :time_eventX              .
}
```

These statements specify that the event is of a certain type (nwr:LeavingAnOrganization), that it involves a entity playing the role of an employer (:companyA) and two entities playing the role of employees (:employeeA,:employeeB), and that it occurred at a certain time (:time_eventX).

A "nwr:LeavingAnOrganization" event will in turn trigger the instantiation of two situations, one preceding the event (:obj-graph-pre-situation-eventX) and one following the event (:obj-graph-post-situation-eventX):

```
:obj-graph-eventX {
    :eventX
        nwr:hasPreSituation        :obj-graph-pre-situation-eventX  ;
        nwr:hasPostSituation       :obj-graph-post-situation-eventX .
}
```

As previously mentioned, each of these situations will correspond to a name graph containing assertions holding in them. In particular, for the example considered we will instantiate the following two named graphs:

```
:obj-graph-pre-situation-eventX {
    :companyA   nwr:employ    :employeeA    ;
                nwr:employ    :employeeB    .
}

:obj-graph-post-situation-eventX {
    :companyA   nwr:notEmploy :employeeA    ;
                nwr:notEmploy :employeeB    .
}
```

stating that before the firing event, both employeeA and employeeB were employed at the company, while after the firing event none of them was working for the company.

Additional assertions may be attached to situation named graphs. These assertions may be used to characterize the time span of the situation, or the provenance of the statements defined in the situation. For instance, the assertions

```
:instances {
    :obj-graph-pre-situation-eventX
         a               nwr:Situation                         ;
         nwr:hasTime      :obj-graph-pre-situation-eventX-time ;
         nwr:producedBy   nwr:reasoner                         .
    :obj-graph-post-situation-eventX
         a               nwr:Situation                         ;
         nwr:hasTime      :obj-graph-post-situation-eventX-time;
         nwr:producedBy   nwr:reasoner                         .
    :obj-graph-pre-situation-eventX-time
         a               time:Interval   ;
         time:hasEnd      :time_eventX    .
    :obj-graph-post-situation-eventX-time
```

```
        a               time:Interval   ;
        time:hasBeg     :time_eventX    .
}
```

permit to assert that the two situations were instantiated by the agent nwr:reasoner, that obj-graph-pre-situation-eventX was in place before eventX, and that obj-graph-post-situation-eventX is in place after eventX. Likewise, we will be able to distinguish situations that are explicitly described in the text and claimed by the sources from situations that are indirectly derived through the nwr:reasoner. In the former case, the named graph has an nwr:attributedTo property with the source, and in the latter case the nwr:producedBy property to the reasoner.

In order to enable expressing events, situations, and to define the conditions and modalities on how to trigger such situations starting from events, the NewsReader Domain Ontology has to fulfil some requirements:

- define the core classes (e.g., Event, Situation) and the basic properties that enable relating them (e.g., to state that a Situation S is a pre-situation of an event E);
- define the type of events that are relevant for the NewsReader use cases, potentially abstracting form the specific way an event is mentioned in the text, so that different variants of the same event (e.g., firing, sacking) can be treated the same way;
- organize events into a taxonomy so to exploit the inferencing capabilities on the subclass relation between events (i.e., if an event triggers some situations, every event more specific than it should trigger the same situations);
- define how situations are triggered by events, specifying which assertions to instantiate in each situation.

4.1.2 Core classes and properties of the NewsReader Domain Ontology

The NewsReader Domain Ontology contains five core classes, which are further specialized in subclasses:

Event : this class is the root of the taxonomy of (proper) event types considered in NewsReader. Any event detected in a text will be an instance of some class of this taxonomy;

DynamicEvent : this is a subclass of Event (for which dynamic changes are defined) that apply to FrameNet frames that can be considered as proper events (e.g., fn:firing);

StaticEvent : this is another subclass of Event for "static" event types considered in NewsReader and which capture more static circumstances (e.g., fn:possession, fn:organization); they typically directly trigger a situation holding at the time the event occurs (a "during situation", differently from pre/post-situations in proper events); a "static" event detected in a text will be an instance of some class of this taxonomy;

Situation : the individuals of this class are actual pre/post/during situations that will be instantiated starting from the event instances detected in the text;

SituationRule : the individuals of this class enable to encode the rules for instantiating pre/post/during situations when a certain type of event is detected;

SituationRuleAssertions : the individuals of this class enable to encode the assertion that has to be instantiated within each pre/post/during situation associated to some event.

Analogously to FrameNet frame elements for frames, the NewsReader Domain Ontology enables to represent the role of an entity in an event. Roles are formalized as object properties: this way, an event instance :eventX can be related to an entity :entityZ participating in it with assertions of the form:

:eventX nwr:hasRoleY :entityZ

where nwr:hasRoleY specify the role of :entityZ in :eventX. Each object property defining a role in the NewsReader Domain Ontology is defined as subproperty of the top object property nwr:hasRole: this way, given any event, we can retrieve the entities participating in it by looking at assertions having as predicate the property nwr:hasRole.

Additional object properties are defined to enable:

- relating an event instance with the actual pre/post/during situations it triggers (resp., object property nwr:hasPreSituation, nwr:hasPostSituation, and nwr:hasDuringSituation);
- relating an event type with the pre/post/during situation rules that should be triggered when an instance of that event type is detected (resp. nwr:triggersPreSituation, nwr:triggersPostSituation, and nwr:triggersDuringSituation);
- relating a situation rule with the assertions that should be instantiated within the situation named graph associated with the rule (resp., nwr:hasSituationRuleAssertion).

Finally, the NewsReader Domain Ontology specifies the properties that can be used as predicate in assertions within a situation named graph. Two typologies of properties are considered:

binary properties : these properties are modelled as object properties and they enable to relate two entities (e.g., see property "nwr:employ" and "nwr:notEmploy" in the situations instantiated for the firing event example previous considered);

unary properties : these properties are modelled as datatype properties and they enable to express facts such as that an entity exists. Typically, the range of such properties is a boolean value type.

For binary properties, whenever appropriate, we defined additional properties characteristics. In particular, two important characterization are in-place:

disjoint properties : two binary properties p, q are defined as disjoint if no individual a can be connected to an individual b by both triples $a\ p\ b$ and $a\ q\ b$.

inverse properties : if two binary properties p, q are defined as one the inverse of the other, an assertion $a\ p\ b$ implies also the assertion $b\ q\ a$, and viceversa.

For instance, in the NewsReader Domain Ontology we defined "nwr:employ" and "nwr:notEmploy" as disjoint (only one of the two can hold at a certain time), as well as "nwr:employ" and "nwr:employedAt" as inverse properties (if :companyA nwr:employ :employeeB, then :employeeB nwr:employedAt :companyA holds, and viceversa).

4.1.3 Formalization of the rules for instantiating situations from events

The formalization of the rules for instantiating situations from events consists in defining the assertions to be instantiated in pre/post/during situations of an event, based on the roles of the entities involved in it. We rely on a two level schema: first, we define for each event type the kind of situations they have to trigger (i.e., whether pre/post/during situations); then, for each situation triggered by an event, we formalize the type of assertions that have to be instantiated, specifying how the roles of the event triggering the situation map to the assertions' subject and object. We illustrate this with a concrete example, based on the event type "ChangeOfPossession", which refers to the event when something (role "possession-theme") passes from an entity (role "possession-owner_1") to another entity (role "possession-owner_2"). An event of type "ChangeOfPossession" has to trigger a pre-situation and a post-situation, each of them asserting some possession statements. To model the relation between an event type and the type of situations it triggers we rely on owl:hasValue restrictions:

```
nwr:ChangeOfPossession   rdfs:subClassOf [
a owl:Restriction ;
    owl:hasValue   nwr:pre_ChangeOfPossession   ;
    owl:onProperty nwr:triggersPreSituationRule ] .

nwr:ChangeOfPossession   rdfs:subClassOf [
    a owl:Restriction ;
    owl:hasValue   nwr:post_ChangeOfPossession   ;
    owl:onProperty nwr:triggersPostSituationRule ] .

nwr:pre_ChangeOfPossession    a nwr:SituationRule .
nwr:post_ChangeOfPossession   a nwr:SituationRule .
```

Note that, by defining the "rule" for instantiating situations based on owl:hasValue restrictions, we can later exploit reasoning to infer that the same pre/post/during situations have to be triggered for any event type more specific than the considered one:

e.g., if we are considering an event of type nwr:Getting, and nwr:Getting is a subclass of nwr:ChangeOfPossession, the same rules for situations defined for nwr:ChangeOfPossession automatically apply also for nwr:Getting, without having to redefine them.

Each nwr:SituationRule individual is then specialized to define exactly how the triples inside the Situation named graph has to be defined. This is done by defining an individual (of type SituationRuleAssertion) for each assertion to be created, having three annotation properties assertions:

nwr:hasSituationAssertionSubject : the object of this triple is the role of the event to be used as subject in the assertion;

nwr:hasSituationAssertionProperty : the object of this triple is the predicate to be used in the assertion. It is either a binary property or an unary property;

nwr:hasSituationAssertionObject : the object of this triple is the role of the event or the data value (in case of unary properties) to be used as object in the assertion.

Consider for instance the nwr:pre_ChangeOfPossession situation rule:

```
nwr:pre_ChangeOfPossession
    nwr:hasSituationRuleAssertion    pre_ChangeOfPossession_assertion1;
    nwr:hasSituationRuleAssertion    pre_ChangeOfPossession_assertion2.
```

This rule triggers the instantiation of two assertions, nwr:pre_ChangeOfPossession_assertion1 and nwr:pre_ChangeOfPossession_assertion2, defined as follows:

```
nwr:pre_ChangeOfPossession_assertion1
    nwr:hasSituationAssertionSubject     nwr:possession-owner_1;
    nwr:hasSituationAssertionProperty    nwr:possess;
    hasSituationAssertionObject          nwr:possession-theme.

nwr:pre_ChangeOfPossession_assertion2
    nwr:hasSituationAssertionSubject     nwr:possession-owner_2;
    nwr:hasSituationAssertionProperty    nwr:notPossess;
    hasSituationAssertionObject          nwr:possession-theme.
```

Therefore, from an event instance :eventX of type nwr:ChangeOfPossession, having roles :instanceX (nwr:possession-owner_1 role), :instanceY (nwr:possession-owner_2 role), and :instanceZ (nwr:possession-theme role), by interpreting the aforementioned rule schema we can instantiate a pre-situation named graph, :eventX_pre, defined as follow:

```
:eventX_pre {
    :instanceX    nwr:possess       :instanceZ .
    :instanceY    nwr:notPossess    :instanceZ .
}
```

where the first assertion is created due to nwr:pre_ChangeOfPossession_assertion1, while the second assertion is due to nwr:pre_ChangeOfPossession_assertion2.

4.1.4 Mappings from external resources to the NewsReader Domain Ontology

A key ingredient of the NewsReader Domain Ontology is the mapping of the FrameNet frames and frame elements to the event types and roles that we defined. This mapping is necessary to translate the annotations provided by the SRL module to our ontology vocabulary, exploited by the reasoning module to instantiate situations from events.

For each event type (modelled as class in the NewsReader Domain Ontology) and each role (modelled as object property in the NewsReader Domain Ontology) we defined some annotations (nwr:correspondsToFrameNetFrame and nwr:correspondsToFrameNetElement) representing the corresponding frames and frame elements. For instance, the following assertions via property nwr:correspondsToFrameNet-Frame are defined for the event type nwr:Giving:

```
nwr:Giving    nwr:correspondsToFrameNetFrame    fn:Giving,fn:Sending,fn:Supply.
```

meaning that, if a frame of type fn:Giving, fn:Sending, or fn:Supply is identified in the text, it has to be considered as an event of type nwr:Giving, and therefore pre/post/during situation rules defined for nwr:Giving should be triggered.

Similarly, given the role nwr:possession-owner_1, it is mapped to the following frame elements with the nwr:correspondsToFrameNetElement assertions:

```
nwr:Giving   nwr:correspondsToFrameNetElement   fn:Seller,fn:Supplier,fn:Lender,
                                                fn:Sender,fn:Donor,fn:Source,
                                                fn:Agent,fn:Exporter,fn:Victim.
```

We also defined the mapping from the NewsReader Domain Ontology event types to SUMO[7] classes (as explained in section 4.2), via nwr:correspondsToSUMOClass annotation assertions. E.g., the following mapping of nwr:Giving to a SUMO class was defined:

```
nwr:Giving    nwr:correspondsToFrameNetFrame    fn:Giving,fn:Sending,fn:Supply.
```

4.2 Building the Event and Situation Ontology version 1

As a first step in building a domain specific ontology, we carried out a statistical analysis of the events in the car data set. We chose to include only events related to FrameNet for this analysis as the frames associated to predicates provide a set of roles (Frame entities); both are needed to formulate the pre and post conditions of the events. We extracted all predicates with an external reference to FrameNet from 65,540 NAF files. This yielded a total of 3,612,511 predicates, 2,147 unique combinations of a lexical unit and a FrameNet frame and 428 unique frames. Note that a frame can be linked to multiple lexical units. In order to select the domain events and related frames, we annotated all predicates as being either contextual, grammatical, cognitive, perceptive or related to communication:

- Communication: all predicates related to communication, communicative gestures, motions and actions: (*remark, write, hush, forbid, howl, smile, censure, translate, nod, sing, wave*)

[7] `http://www.ontologyportal.org`

Predicate type	Number of frames	Unique predicates	Total predicate frequency
Communication	88	396	818,291 (22.65%)
Cognitive	36	222	337,766 (9.34%)
Perception	9	50	96,821(2.68%)
Grammatical	78	173	1,002,109 (27.73%)
Contextual	234	1306	1,357,524 (37.57%)
Totals	445	2147	3,612,511

Table 5: Statistics on the predicates related to a FrameNet frame per predicate type

- Cognitive predicates: all predicates expressing states of mind and mental processes that may or may not induce actions: (*prefer, expect, worry, hope, deduce, classify, interpret, know, adopt, choose*);

- Perception: all predicates that denote physical experiences and sensations: (*feel, sense, hurt, observe, find, spy, taste*);

- Grammatical: all predicates that express aspect of another verb (*stop, begin, continue*) and light verbs: (*prevent, stop, take, remain, precede, engage, contain, enlarge, imply, emerge, achieve*);

- Contextual predicates. All predicates that do not belong to one of the previous classes are contextual and potentially important for the domain: (*fluctuate, meet, break, melt, buy, accompany, refresh, sleep*).

All predicates belonging to Communication, Cognition and Perception will be used for the attribution model, whereas the grammatical predicates are ignored for the time being because they do not introduce events in a timeline but rather express properties of events. The contextual predicates then form the group of potential important events for the car domain. Table 5 shows the statistics on the extracted predicates related to a FrameNet frame. About 63% of all predicates found is not domain-specific; grammatical and communication related predicates make up the majority of the not domain-specific predicates with 27.73% and 22.65% respectively. The contextual predicates dominate the statistics, both in the number of unique frames (234), unique predicates (1306) and total predicate frequency (1,357,524).

For building the ontology, we need to define the following structures:

1. A hierarchy of events that are important for the domain and allow for inferencing;

2. A set of properties that allows for defining the most salient pre and post conditions of the event;

3. A set of statements that define the roles of the entities affected by the change.

4.2.1 Hierarchy of events

To derive the first component for the ontology, a hierarchy of important domain events, we used the list of extracted contextual predicates with FrameNet mappings. A complete overview of all these frames is given in the appendix 9. As such, we started with 234 frames and 1,306 unique predicates with potentiality to be domain important. To scope this set, we put a threshold on the frames: all frames that were found only once, and in combination with a predicate with a frequency under 100, were not taken into account. As a result, 183 frames remained. Next, we experimented with three approaches to select a set of frames for modeling the event ontology.

In the first approach, we tried to select the most important frames by sorting on: a) the number of unique predicates that were found for this frame; b) the frequency of these predicates in our data; c) a combination of both. However, it turned out that these frequency statistics were not reliable enough. The number of predicates found for a frame depends solely on how many predicates have been defined in FrameNet. As such, it is not a strong pointer to dominant concepts. Additionally, some predicates that are known to be high frequent: this biases the frequency statistics we derived; a predicate such as *make* sometimes makes up half of the total predicate frequency of a frame.

In the second approach, we experimented with manually relating the frames from the car data back to the FrameNet to see if we could conceptually group and select concepts for the ontology. This turned out to be problematic as well, since there is no full subclass hierarchy in FrameNet. Also, the frames themselves are organized by frame-semantic principles, meaning that some frames group lexical units that represent different concepts from a more ontological point of view. For instance fn:Forming_Relationships groups both *marry* and *divorce* and fn:Change_position_on_a_scale encompasses *increase* and *decrease*. As such, we decided to use FrameNet in a later stage of modeling the ontology.

In the third and final approach, we turned to another background model to organize the frames. For this, we have used the SUMO ontology[8] as it is freely available, well-documented, it has a good coverage and is mapped to English Wordnet. First, we made a selection of the 183 frames based on their expected importance for the domain: frames such as Cooking_creation, Ingest_substance and Location_of_light were left out. This resulted in 92 frames with the potential to be domain specific. The workflow for defining the hierarchy of dynamic event classes is as follows:

1. The initial and unstructured set of 92 frames was mapped manually to SUMO classes by means of subclass and equivalence relations. All frames that expressed static events were set aside.

2. From this mapping, we selected four top nodes in SUMO that represented the main conceptual clusters for the frames expressing dynamic events: Motion, InternalChange, ChangeOfPossession and IntentionalProcess. In this step, we also started to group similar frames into one class. For instance, the main difference between the

[8] `www.ontologyportal.org`

frames Departing and Quitting_a_place is a specification of the entity that moves. For our purposes, this level of granularity is not necessary. As such, both frames have been defined as corresponding to the ESO class Departing.

3. Next, we checked the SUMO class hierarchy of Motion, InternalChange, ChangeOfPossession and IntentionalProcess to select additional classes that may be of importance for the car domain, such as Investing and Importing.

4. We defined four hierarchies consisting of ESO classes with a mapping to SUMO and FrameNet and potential ESO classes with only a SUMO mapping.

5. To increase the coverage, we mapped back from these ESO classes to FrameNet frames. For this, we used the existing frame-to-frame relations in FrameNet (Ruppenhofer *et al.*, 2006). These additional were either a) found in the car data, but previously ruled out by the thresholds or b) not found in the car data but a frame for the ESO class does exist in FrameNet. In some cases, frames were found for which we had no SUMO-based ESO class. In those cases, a new ESO subclass was defined. Also, for some SUMO-based ESO classes no corresponding frame could be found. These classes were kept in the ontology nonevertheless as placeholder for future extensions. As such, we have ESO classes with mappings to both FrameNet and SUMO, ESO classes with only a mapping to FrameNet, and ESO classes with only a mapping to SUMO. Furthermore, to keep the hierarchy clean, we opted to use single inheritance only for all event classes in the ontology.

For the static events that were set aside earlier, we performed the same workflow. However, the static events are represented as a flat hierarchy in this version of the ontology since we have found only few frames for static events in our data to define a hierarchy.

4.2.2 Properties for defining pre and post conditions of an event

The second component of the ontology consists of pre and post conditions that state which situation holds before and after an event. For instance, for events pertaining to the class Translocation, the property 'atPlace' defines that some entity is at some location before a translocation event (pre condition) and the entity is at another location after the event (post condition). All condition properties were hand-built based on the shared semantics of the predicates related to a frame.

For some properties that express pre and post conditions, we expect that they are either too strong or too weak to capture the implications of an event. For instance, the property 'possess' is used for situations where someone actually possesses something (nwr:Buying) but also for event classes where this possession seems weakened from 'ownership' into some sort of 'having', e.g. nwr:Import and nwr:Taking. From the results of the reasoner, we can finetune these implications if necessary.

For static events, we defined conditions that are true for the duration of the static events. For the static event 'InEmployment' the property 'employedAt' defines that some

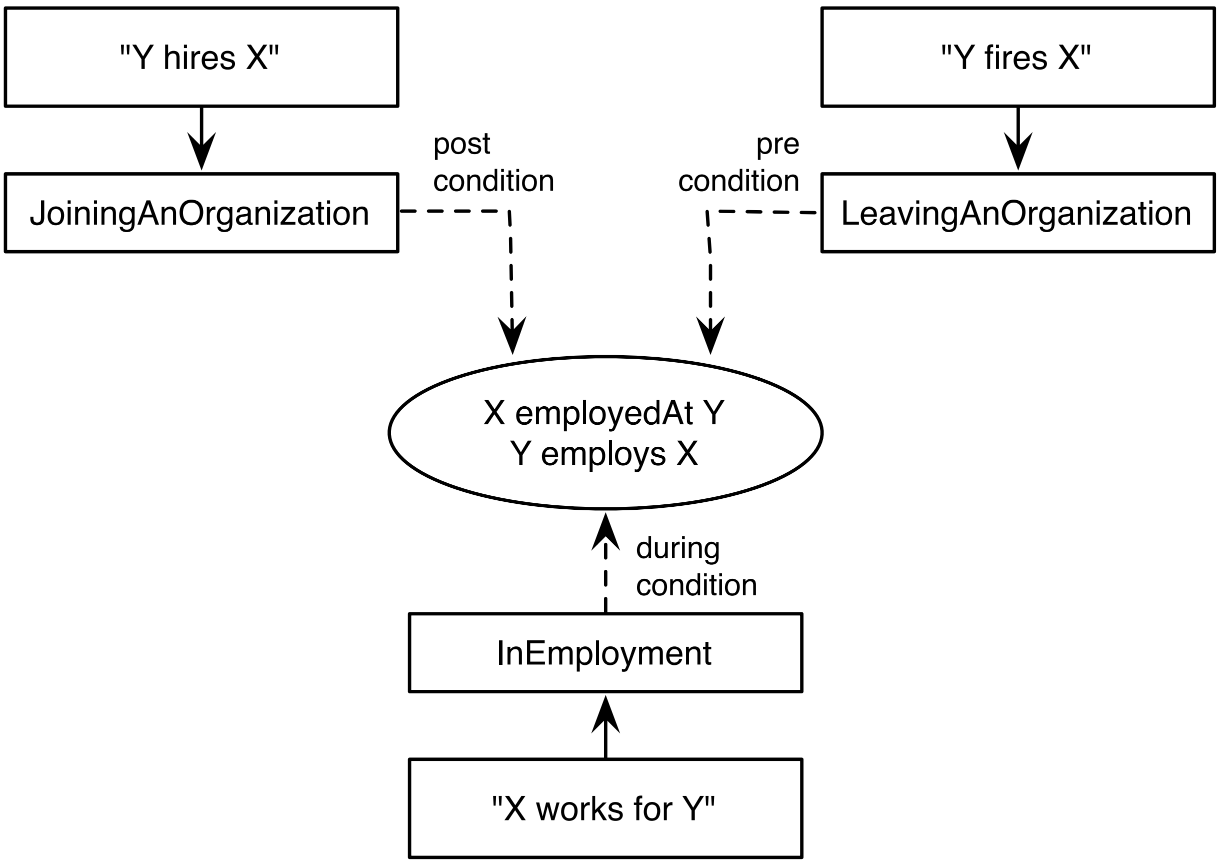

Figure 2: Overview of the dynamic event class hierarchy.

person is employed at some employer. As illustrated in figure 2, the same property is used as pre condition for the dynamic event class LeavingAnOrganization and as post condition for the dynamic event class JoiningAnOrganization. Where applicable, related dynamic and static events have been related as such by means of a shared condition property. As a result, the relation between an inferred condition of a dynamic event and the more explicit condition of a static events becomes more easily understandable.

4.2.3 Roles for the entities affected by an event

For the roles of the entities that are affected by the change, we used a selection of FrameNet Frame Entities (FEs), derived from the mappings of an ESO class to FrameNet frames. As such, we define which roles are important for modeling our domain and which roles are not. For the class Translocation, the entity that translocates maps to FrameNet FE 'Self_mover', 'Theme' and 'Driver' and the entity that expresses the location to 'Location' and 'Goal'. Other FEs that are specified in translocation frames such as 'Manner', 'Distance' and 'Speed' do not play a role in formulating the salient conditions of an event. As such, they are not incorporated.

4.3 Event and Situation Ontology version 1

In this section, we first give statistics of the content of the current version of the ontology and provide an overview of the current contents of the ontology. The OWL version of the ontology can be downloaded from the NewsReader website.

The first version of the ESO consists now of 59 event classes divided over dynamic events (50) and static events (9). The dynamic event class hierarchy consists of four major nodes: ChangeOfPossession (16 subclasses), Motion (10 subclasses), InternalChange (11 subclasses) and IntentionalEvents (11 subclasses). An overview of the dynamic event hierarchy is presented in Figure 3. The static events are not modeled into a hierarchy; an overview of the static events is presented in table 6.

For 53 classes, we have one or multiple mappings to FrameNet frames. In total, 94 mappings to FrameNet were made, covering 532 unique combinations of a predicate and a frame which equals to a total frequency of 990,152 predicates related to this ontology.

Additionally, 49 out of 59 event classes have a mapping to SUMO. Currently, only equivalence mappings are incorporated. An overview of all mappings from ESO classes to SUMO and FrameNet can be found in table 10.

We defined 30 situation rules that express the conditions for dynamic and static event classes, e.g. 'pre_Translocation' and 'during_BeingAtAPlace'. These situation rules cover 35 out of 50 dynamic event classes, and all 9 static event classes. Note that some situation rules are defined at the top of a class tree and are thus inherited by all its subclasses. The situation rules for dynamic and static events are presented in table 7.

These 30 different situation rules trigger in total 41 unique situation rule assertions where the implications of a change or state are expressed, e.g. 'pre_Translocation_assertion_1'. Note that one rule can be coupled with one or more assertions.

In the assertions, we use properties that define the relation between the roles of the entities that are affected by a dynamic or static event. Currently, we defined 24 properties (20 binary and 4 unary) such as 'atPlace', 'employedAt' and 'hasInPossession'. An overview of these properties is shown in table 8.

Finally, we defined 33 different roles for the entities such as 'atPlace-theme', employment-employee' and 'possession-owner'. An overview of the roles is presented in table 9.

Each role is mapped to one or more Frame Entities in FrameNet. All FrameNet Entities we currently use are listed in table 11.

ESO Class	SUMO mapping	FrameNet mapping
Arriving	Translocation	Arriving
		Vehicle_landing
Attacking	Attack	Attack
BeingAtAPlace	-	Being_located
		Presence

		Residence
		Temporary_stay
BeingInAPersonalRelationship	-	Personal_relationship
BeingInExistence	-	Existence
BeingLeader	-	Leadership
BeingOperational	-	Being_operational
Borrowing	Borrowing Borrowing	
FinancialTransaction	Buying	Commerce_buy
ChangeOfLeadership	-	Change_of_leadership
ChangeOfPossession	ChangeOfPossession	-
ChangeOfRelationship	-	Forming_relationships
ChangePositionInOrganization	TransferringPosition	-
ChangingShape	ShapeChange	Manipulate_into_shape
		Reshaping
Collaboration	-	Collaboration
Constructing	Constructing	Building
	Making	
Creating	Creation	Creating
		Intentionally_create
Damaging	Damaging	Damaging
		Render_nonfunctional
Destroying	Destruction	Cause_to_fragment
		Destroying
Distribution	-	Dispersal
Escaping	Escaping	Escaping
		Fleeing
Exporting	Exporting	Exporting
FinancialTransaction	FinancialTransaction	Commercial_transaction
Getting	Getting	Getting
		Receiving
Giving	Giving	Giving
		Sending
		Supply
HavingInPossession	-	Possession
Importing	Exporting	Importing
InEmployment	-	Being_employed
		Employing
Injuring	Injuring	Cause_harm
		Experience_bodily_harm
Installing	Installing	Installing
IntentionalEvent	IntentionalProcess	Intentionally_act
InternalChange	InternalChange	-
Investing	Investing	-
JoiningAnOrganization	JoiningAnOrganization	Get_a_job
		Hiring

Killing	Killing	Execution
		Killing
Leaving	Leaving	Departing
		Quitting_a_place
		Setting_out
		Vehicle_departure
LeavingAnOrganization	LeavingAnOrganization	Firing
		Quitting
Lending	Lending	Lending
Manufacturing	Manufacture	Manufacturing
Meeting	Meeting	Assemble
		Come_together
		Social_event
Merging	Combining	Amalgamation
		Cause_to_amalgamate
Motion	Motion	Motion
OrganizationalEvent	OrganizationalEvent	-
Paying	Payment	Commerce_pay
Placing	Putting	Placing
QuantityChange	QuantityChange	Cause_change_of_position_on_a_scale
		Cause_expansion
		Cause_proliferation_in_number
		Change_of_quantity_of_possession
		Change_position_on_a_scale
		Expansion
		Proliferating_in_number
Removing	Removing	Removing
Renting	Renting	Renting
RentingOut	-	Renting_out
Replacing	Substituting	Replacing
		Take_place_of
Selling	Selling	Commerce_sell
Separating	Separating	Becoming_separated
		Separating
SocialInteraction	SocialInteraction	-
StaticEvent	-	State
Stealing	Stealing	Theft
Taking	UnilateralGetting	Taking
Translocation	Translocation	Cause_motion
		Cotheme
		Intentional_traversing
		Operate_vehicle
		Ride_vehicle
		Self_motion
		Travel

		Traversing
		Use_vehicle
Transportation	Transportation	Delivery
		Bringing
Working	-	Work

Table 10: Mappings from ESO classes to SUMO classes and FrameNet frames

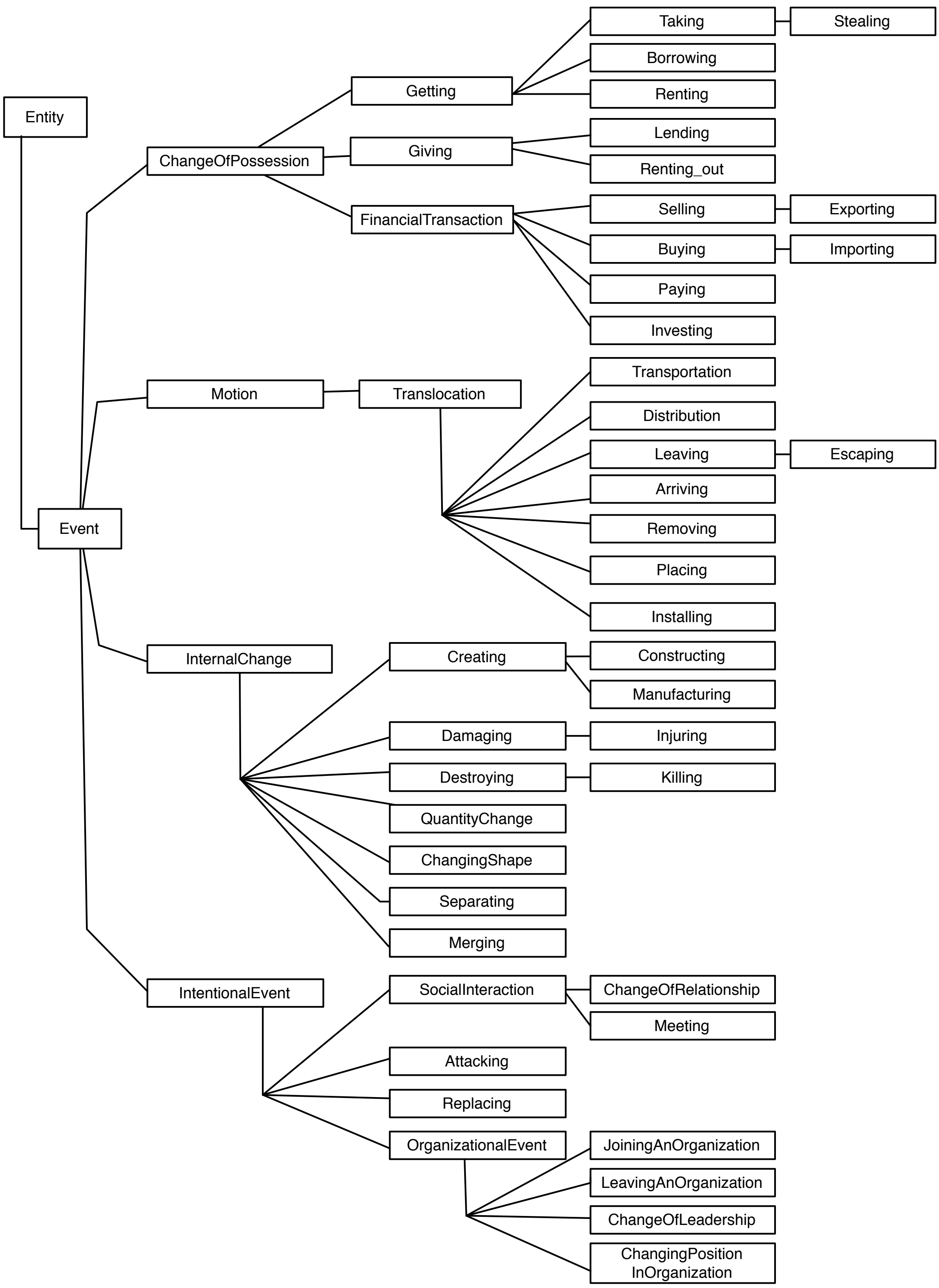

Figure 3: Overview of the dynamic event class hierarchy

Static events
BeingAtAPlace
BeingInAPersonalRelationship
BeingInExistence
BeingLeader
BeingOperational
Collaboration
HavingInPossession
InEmployment
Working

Table 6: Overview of static events

Situation rules static events	Situation rules dynamic events
during-BeingAtAPlace	pre/post_ChangePossession
during-BeingInAPersonalRelationship	pre/post_Creating
during-BeingInExistence	pre/post_Destroying
during-BeingLeader	pre/post_Installing
during-BeingOperational	pre/post_JoiningAnOrganization
during-Collaboration	pre/post_LeavingAnOrganization
during-HavingInPossession	pre/post_Merging
during-InEmployment	pre/post_QuantityChange
during-Working	pre/post_Separating
	pre/post_Translocation

Table 7: Overview of situation rules for static and dynamic events

Binary properties		Unary properties
atPlace	hasInPossession	exist
isPlaceOf	inPossessionOf	notExist
notAtPlace	notHasInPossession	isOperational
notIsPlaceOf	notInPossessionOf	notIsOperational
collaboratesWith	inRelationshipWith	
notCollaboratesWith	notInRelationshipWith	
employedAt	hasLeader	
employs	isLeaderOf	
notEmployedAt	isValueOf	
notEmploys	hasValue	

Table 8: Overview of Unary and Binary properties

Roles		
atPlace-location	leader-entity	quantity-value_1
atPlace-theme	leader-governed_entity	quantity_value_2
collaboration-partner_1	merging-theme_1	relationship-partner_1
collaboration-partner_2	merging-theme_2	relationship-partner_2
creating-theme	merging-theme_3	separating-theme_1
destroying-theme	operational-theme	separating_theme_2
employment-employee	possession-owner	separating_theme_3
employment-employer	possession-owner_1	translocation-goal
exist-theme	possession-owner_2	translocation-source
installing-goal	possession-theme	translocation-theme
installing-theme	quantity-theme	working-entity

Table 9: Overview of roles

Frame Entities			
Agent	Executed	Lessee	Sender
Borrower	Exporter	Location	Set
Buyer	Final_number	Owner	Source
Carrier	Final_value	Part_1	Supplier
Component	Fixed_location	Part_2	Theme
Cotheme	Goal	Partner_1	Traveler
Created_entity	Governed	Partner_2	Undergoer
Deliverer	Guest	Perpetrator	Value_1
Device	Individuals	Possession	Value_2
Donor	Initial_number	Product	Vehicle
Driver	Initial_size	Recipient	Victim
Employee	Initial_value	Resident	Whole
Employer	Item	Result_size	Whole_patient
Entity	Leader	Self_mover	
Escapee	Lender	Seller	

Table 11: FrameNet Frame Entities reused for specifying the roles of entities

5 Vocabularies

Following best practices in Semantic Web technologies, ESO reuses parts of two existing vocabularies: there are mappings from our ontology to FrameNet on class and role level, and mappings to SUMO on class level. As such, we can define our classes specific for the car domain, without adhering to modeling choices in FrameNet and SUMO. Through these mappings, ESO serves as a hub to other vocabularies as well, as is illustrated by figure 4. The boxes represent the different vocabularies: ESO, SUMO, FrameNet, Princeton Wordnet (PWN) and a selection of Wordnets in the Global Wordnet grid (Italian Wordnet, Dutch Wordnet and Spanish Wordnet). The coloured circle represents an ESO class and its mappings in the different vocabularies which each have their own modeling principles.

The ESO provides a domain-specific hierarchy of events and their implications across languages. For instance, the output of the Semantic Role Labeling and the Predicate Matrix assigns a predicate 'departs' to the FrameNet frame Departing. This frame corresponds to the ESO class Departing where a) super and subclasses and b) implications are defined. From the output of the WSD module, we know that the predicate 'departs' relates to a Princeton Wordnet (PWN) synset ID. This PWN synset has pointers to Wordnets in the Global Wordnet Grid, where Wordnets are represented for several languages, including Dutch, Italian and Basque. From this mapping, we can infer that the initial predicate 'departure' is equivalent to the Italian synset 'partire' and the Dutch synset 'vertrekken' and that the implications in ESO for the class Departure also holds for these synsets and their hyponyms. The cross-lingual mapping provides the basis for cross-lingual event extraction and modeling as well.

The PredicateMatrix[9] and the GlobalWordnetGrid[10] can be used to get coverage of large vocabularies and across many different languages to apply the model to texts.

[9] `http://adimen.si.ehu.es/web/PredicateMatrix`
[10] `www.globalwordnet.org`

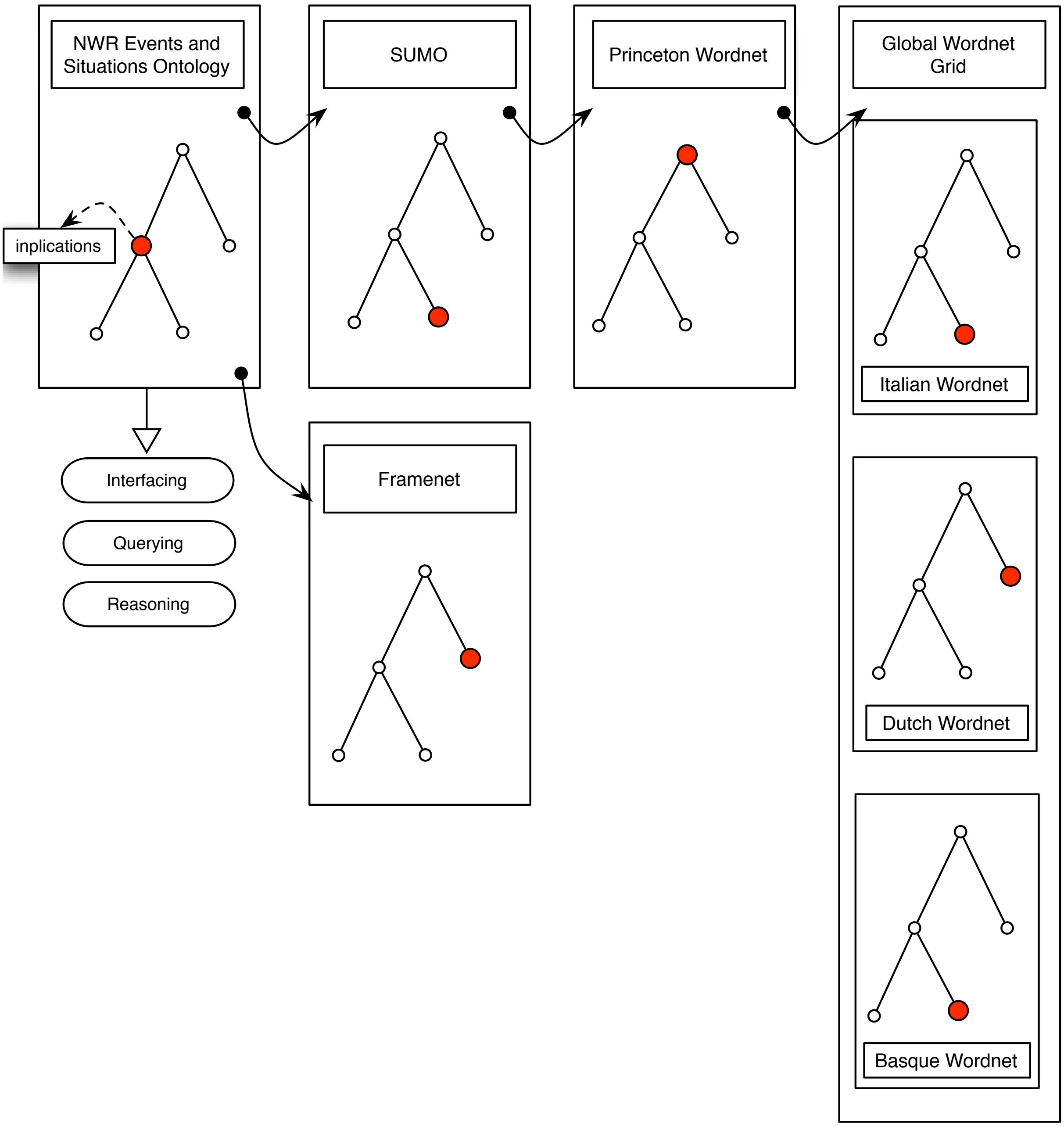

Figure 4: NewsReader Events and Situations ontology and direct and indirect mappings to other vocabularies

6 Integrating ESO into the Predicate Matrix

The Predicate Matrix allows the interoperability between several sources of predicate information. As SemLink (Palmer, 2009), the Predicate Matrix merges several models of predicates such as VerbNet (Kipper *et al.*, 2000), FrameNet (Baker *et al.*, 1998), PropBank (Palmer *et al.*, 2005) and WordNet (Fellbaum, 1998). The Predicate Matrix also contains for each predicate features of the ontologies integrated in the Multilingual Central Repository (Gonzalez-Agirre *et al.*, 2012) like SUMO (Niles and Pease, 2001), Top Ontology (Álvez *et al.*, 2008) or WordNet domains (Bentivogli *et al.*, 2004). The mappings between such knowledge bases allow to take advantage from their individual strengths. For example, the coverage of PropBank or the semantic relations among events and participants of FrameNet. Moreover, it is also possible to automatically integrate all knowledge connected to any of its components, as in the case of ESO.

Both FrameNet and SUMO labels integrated in ESO are used to connect ESO to the PM. For example, the predicate **sell.01** of PropBank belongs, according to their mappings in the PM, to the frame *Commerce_sell* of FrameNet. Thus, this predicate and its arguments would be mapped to ESO as shows table 12. Moreover, the frame can also be linked through the SUMO classes. For instance, the predicate **drain.01** of PropBank belongs to the frame *Emptying* that is not considered in ESO. However, it also belongs to the class *Removing* of SUMO and, in consequence, the mappings in table 13 can be obtained. In this way, ESO is connected to **2235** predicates and **3445** different roles of the PM.

PB-pred	PB-arg	FN-frame	FN-fe	ESO-class	ESO-role
sell.01	arg_0	Commerce_sell	Seller	Selling	possession-owner_1
sell.01	arg_1	Commerce_sell	Goods	Selling	possession-theme
sell.01	arg_2	Commerce_sell	Buyer	Selling	possession-owner_2

Table 12: Mapping between PropBank and ESO through FN.

PB-pred	PB-arg	SUMO-class	FN-fe	ESO-class	ESO-role
drain.01	arg_0	Removing	Theme	Removing	translocation-theme
drain.01	arg_1	Removing	Source	Removing	translocation-source

Table 13: Mapping between PropBank and ESO through SUMO.

6.1 SRL post-processing based on ESO

The ESO ontology defines the semantics of the FrameNet frames in the car domain, and is also employed for filtering the frames assigned by the SRL module. For instance, the SRL module identifies three FrameNet frames for the predicate *fires*: Firing, Shoot_projectiles and Use_firearm. The desired frame for the car domain is Firing, which pertains to predicates that express the discharge of employees. The ESO ontology contains the ESO class "LeavingAnOrganization" which has explicit mappings to the FrameNet frames "Firing" and "Quitting". Hence, when the SRL module comes up with an ESO class for a predicate, this can be used for filtering of the FrameNet frames which do not correspond to that class. This is exactly what is done in the SRL post-processing phase. In a similar manner the

FrameNet role elements of the SRL module are assessed. If the FrameNet entity of the role corresponds to an ESO class and the role element corresponds to an ESO object property, then the FrameNet role is confirmed as correct; otherwise the FrameNet element is being marked as incorrect.

A quantitative analysis was performed on the 1.3 million articles from the car domain to assess the filtering by the post-processing module. The 30 ESO classes that are currently in the Predicate Matrix cover 24.23% of the real data predicates. Furthermore, based on the identified ESO classes, we were able to come up with a judgement on 36.48% of the FrameNet frames, most of which were identified as correct, and the rest as incorrect. The FrameNet frames are present on average almost once per predicate. The success rate with respect to roles is lower; 10.8% of the roles could be identified with ESO roles, which led to 29.22% of judged FrameNet roles. Around half of those FrameNet roles were marked as correct, the rest as incorrect. One of the reasons for the lower success in the case of the roles comes from the fact they are more sparse: there is 0.805 FrameNet elements per role, which is a lower average than in the case of the predicates. Another reason is the mapping of roles to ESO consists of two checks instead of one, as explained before in this subsection.

Analysis	Predicates	Roles
Total:	84.877.181	184.573.548
With ESO:	19.542.339 (24.23 %)	19.940.244 (10.80%)
Total FrameNet:	80.649.990	148.582.438
Positive FN:	17.515.482 (21.72 %)	20.899.385 (14.07 %)
Negative FN:	11.902.592 (14.76 %)	21.023.911 (14.15 %)
Judged FN:	29.418.074 (36.48 %)	41.923.296 (29.22 %)
Unjudged FN:	51.231.916 (63.52 %)	106.659.142 (71.78 %)

Additionally, we carried out a more in-depth and qualitative analysis of the coverage of the Events and Situations Ontology. Such analysis could be perceived as an exploratory phase for future improvements of ESO.

We observed that 41.240.519 (or 48.59 %) of all predicates have neither an ESO, nor FrameNet mapping. As this is a rather significant segment of all the predicates, we investigated which predicate mentions with no such mappings occur most often. According to our analysis, these are: include, work, seat, do, manager, service, hold, and so on. The most often occuring FrameNet frames with no mapping to ESO are Statement, Choosing and Text_creation. These frames were deliberately not incorporated in ESO as they relate to speech acts and not to domain specific predicates. The most frequent frames with mapping to ESO are Possession, Change_position_on_a_scale, Bringing and Arriving. The most useful ESO classes to discriminate on the FrameNet frames are: Getting, Translocation and HavingInPossession. It is worth mentioning, however, that only 30 ESO classes and their corresponding 53 frames from FrameNet were found in the processing phase. This is because the remaining classes were excluded in the processing modules run on the data. The complete output of the analysis can be found on NewsReader's github: `https://github.com/newsreader/eso_stats`.

7 Integration Domain Model

7.1 Domain adaptation of Natural Language Processing

The adaptation of the natural language processing pipeline to the financial economic domain can be split up into two tasks: improving recall and improving precision.

In particular for the named entities, recall can be significantly improved by including external domain resources such as CrunchBase to link entities to. This work is focused on two main tasks:

- Using domain-specific databases and/or resources as gazetteers.
- Using domain-specific unlabelled data to automatically induced word representations.

The inclusion of domain-specific resources such as CrunchBase will allow the use of gazetteer-based features in the training process of Named Entity Recognition and Classification (NERC) models (Agerri *et al.*, 2014). Previous work (Ratinov and Roth, 2009) has shown that NERC is a knowledge intensive process and that gazetteer-based features substantially improve named entity detection and classification. This is particularly important when the gazetteers injected in the learning process are specific to the financial and economic domain.

The second avenue to improve the NERC performance on the financial and economic domain consists of using automatically induced word representations from large amounts of domain-specific unlabeled data. This approach is the subject now of intensive research in the NLP field (Passos *et al.*, 2014; Turian *et al.*, 2010; Ratinov and Roth, 2009) with very good results.

In NewsReader we will focus on word representations based on three different techniques: two based on clustering methods and the third one on distributed representations. Among the clustering methods, we will explore features based on Brown hierarchical clustering (Brown *et al.*, 1992). Brown's algorithm clusters words to maximize the mutual information of bigrams. We will also use Clark's algorithm to clustering words into classes based on morphological and distributional information (Clark, 2003). Both clustering methods are unsupervised and automatically induced the clusters from unlabeled data.

The third method is based on distributed representations of words, which are usually called *word embeddings*. Each word embedding is a vector which can be used to measure the distance between words, namely, how similar one word is to another one. We will use the skip-ngram algorithm of Mikolov *et al.* (2013).

The aim of using such representations is to be able to generalize over the training data to other entities that might be domain-specific. Let us assume that the 'Mercury' brand does not appear in the training data, but that the 'Ford' brand does. Furthermore, let us assume that both words occur in the same cluster. In this case, by using the clusters as features, we can generalize from 'Ford', which does appear in the training data, to 'Mercury' which it does not, because both would occur in the same cluster. This technique allows to bias the learning of NERC models towards the corpus on which the clustering was performed. If that corpus is domain-specific, then we would be, in a way, doing semi-supervised domain adaptation.

However, there will always be entities that are not included in openly available resources. For those (e.g. analysts of companies who are quoted in an article), we are creating a module to induce an entity description from the data. This module would gather all mentions of an entity that cannot be linked to an external resource with a confidence that surpasses a certain threshold and creates a NewsReader entity instance for them which can be enriched with information mentioned around them such as that they are working for a particular company. We will also add a list of common (job) titles in the domain to be able to corefer entity mentions that do and do not include these titles in the coreference module.

To improve the precision of the named entity recognition module, we will add a post-processing step to ensure that the entity mentions that are selected form a complete noun phrase. This can filter out incomplete entity mentions such as "LLC" and "M.". Furthermore, named entity class errors such as organisation/event confusion for the events mentioned in the TechCrunch dataset can be resolved by adding information from the CrunchBase resource. In future versions of the event recognition model, we also aim to cover nominalisations of events.

NED knowledge based-reranking

One of the problems we have encountered is that the named entity disambiguation module does find the correct resource to link an entity to, but because the baseline module is agnostic to the domain, it may not rank the correct resource at the top. We have created a module to try to correct this such that the linking module gives preference to entities that are connected to the automotive and technology domains, which should help link 'Mercury' mentioned in the automotive industry dataset to the car model 'Mercury' instead of the planet.

The module works as follows:

Select named entities the current version of the module only works on named entities for which the NED module has found one or more DBpedia resource

Rerank DBpedia resource based on whether the category to which the resource belongs is central to the automotive industry domain

Obtain final ranking based on string distance since the category hierarchy favours more specific categories, sometimes a resource that is too specific can get selected, by correcting for string distance between the entity mention and the DBpedia resource name this is corrected.

The category hierarchy was induced by manually checking the 500 most commonly occurring named entities with DBpedia resources assigned by the Y1 version of the NewsReader pipeline on the 64,540 automotive industry news articles processed in 2013. From these, the 428 correctly linked entities were selected and for each of these their DBpedia ontology class in the DBpedia 2014 ontology were looked up and ranked by frequency.

From this, we deduced the following 24 classes that are most relevant to the industry domain according to this data sample (from most frequent to least frequent):

- Company
- PopulatedPlace
- Settlement
- Town
- MeanOfTransportation
- TelevisionShow
- Person
- Region
- Village
- PeriodicalLiterature
- FormulaOneTeam
- Building
- Place
- Work
- Legislature
- Organisation
- Plant
- GovernmentAgency
- Film
- OfficeHolder
- Food
- FormulaOneRacer
- Senator

- NationalFootballLeagueEvent

To speed up the module, we generated a resource that contains only DBpedia resource titles that pertain to one of these classes. During runtime, the module then assigns a score to each of the DBpedia resource candidates returned by the NED module based on the class the resource belongs to (Company obtains a score of 24, NationalFootballLeaveEvent a score of 1) minus the Levenshtein distance.[11]

If the candidates can be reranked by the module, it adds an extra element to the NED layer:

```
<entity id="e4" type="MISC">
  <references>
    <!--European-->
    <span>
      <target id="t55"/>
    </span>
  </references>
  <externalReferences>
    <externalRef resource="spotlight_v1" reference="http://dbpedia.org/
       resource/Emigration_from_Europe" confidence="5.313301E-25" reftype="
       en"/>
    <externalRef resource="spotlight_v1" reference="http://dbpedia.org/
       resource/European_Boxing_Union" confidence="1.1661279E-13" reftype="
       en"/>
    <externalRef resource="spotlight_v1" reference="http://dbpedia.org/
       resource/European_Figure_Skating_Championships" confidence="1.6860251
       E-11" reftype="en"/>
    <externalRef resource="spotlight_v1" reference="http://dbpedia.org/
       resource/Culture_of_Europe" confidence="4.737635E-11" reftype="en"/>
    <externalRef resource="spotlight_v1" reference="http://dbpedia.org/
       resource/European_American" confidence="7.080303E-11" reftype="en"/>
    <externalRef resource="spotlight_v1" reference="http://dbpedia.org/
       resource/PGA_European_Tour" confidence="1.1459004E-10" reftype="en"/>
    <externalRef resource="spotlight_v1" reference="http://dbpedia.org/
       resource/UEFA" confidence="1.2734978E-7" reftype="en"/>
    <externalRef resource="spotlight_v1" reference="http://dbpedia.org/
       resource/Ethnic_groups_in_Europe" confidence="4.1297187E-7" reftype="
       en"/>
    <externalRef resource="spotlight_v1" reference="http://dbpedia.org/
       resource/European_Union" confidence="0.002007139" reftype="en"/>
    <externalRef resource="spotlight_v1" reference="http://dbpedia.org/
       resource/Europe" confidence="0.99799234" reftype="en"/>
    <externalRef resource="vua-type-reranker-v1.1" reference="http://dbpedia
       .org/resource/Europe" confidence="19"/>
```

[11] http://en.wikipedia.org/wiki/Levenshtein_distance

```
    </externalReferences>
  </entity>
```

otherwise, the top candidate as defined by the NED module is used by follow-up modules.

Naturally, this first version of the reranking module is limited in its scope as only 24 DBpedia ontology classes are considered, which cover only 1,698,717 of the 12,107,756 DBpedia resources that have a DBpedia ontology class assigned to it.[12] but it is a first step towards filtering out the noisiest named entity disambiguation candidates.

The module can be downloaded at `https://github.com/newsreader/vua-nedtype-reranker`

7.2 Domain adaptation of Event modeling

The generic module for Event modeling described in deliverable Vossen *et al.* (2013) generates a SEM representation for event instances from the mentions in NAF representations of the processed documents. In some cases, the explicit values of the predefined properties are expressed, but in most cases the events are represented that indirectly imply these changes. In the next example , we see a case of explicit expression of a *work-for* relation and the way these relations are expressed in the Semantic Role Layer. There is a reference to a PropBank sense of *work* and the role A0 and A2 for *the graduate engineer* and *VW group* respectively. In this case, we do not find any mappings to FrameNet. In Figure 6, we then see what SEM representation is generated for this event. An instance of a workEvent is created, which is a direct reflection of the Semantic Role structure.

In the next example 7, we see that a dynamic event of *hiring* is represented in the SRL layer by the current module. In this case, a FrameNet mapping has been generated. We can also see in Figure 8 that the corresponding SEM structure is generated with these FrameNet mappings. From a structural point of view, the representations are the same. Both the static as the dynamic event are represented as an event.

The reasoner will read these resulting SEM structures in combination with the ontology described in section 4.3. It will apply different rules for both structures based on the hierarchical mapping of the events and the roles. For the static verb *work*, it applies the following deductions:

1. ProbBank "work.01" maps to the static event class InEmployment, potentially based on a domain adapted PredicateMatrix;

2. The PropBank roles A0 and A1 are mapped to nwr:employment-employee and nwr:employment-employer;

3. The during_InEoplyment_assertion_1 is applied which generates the new triple:

```
nwr:data/cars/inferred-graph1>{
```

[12]Not every DBpedia resource has a DBpedia ontology class assigned to it. See for example `http://dbpedia.org/page/God` which only has YAGO categories assigned to it

Figure 5: SRL: Static event representing a *work-for* relation

```
The graduate engineer has worked for VW group since 1990 in several divisions within production.

   <predicate id="pr10">
     <!--worked-->
     <externalReferences>
       <externalRef reference="work.01" resource="PropBank"/>
     </externalReferences>
     <span>
       <target id="t90"/>
     </span>
     <role id="rl21" semRole="A0">
       <!--The graduate engineer-->
       <span>
         <target id="t86"/><target id="t87"/>
         <target head="yes" id="t88"/>
       </span>
     </role>
     <role id="rl22" semRole="A2">
       <!--for VW group-->
       <span>
         <target head="yes" id="t91"/>
         <target id="t92"/> <target id="t93"/>
       </span>
     </role>
     <role id="rl23" semRole="AM-TMP">
       <!--since 1990 in several divisions within production-->
       <span>
         <target head="yes" id="t94"/>
         <target id="t95"/>... ETC...<target id="t100"/>
       </span>
     </role>
   </predicate>
```

Figure 6: SEM: Static event representing a *work-for* relation

```
<http://www.newsreader-project.eu/data/cars/2009/1/19/4VGW-S1T0-TWSV-40J7.xml#workEvent>
            a               sem:Event , <http://www.newsreader-project.eu/ontologies/propbank/work.01> ;
            rdfs:label      "work" ;
            gaf:denotedBy  <http://www.newsreader-project.eu/data/cars/2009/1/19/4VGW-S1T0-TWSV-40J7.xml#
            char=455,461&word=w90&term=t90> ,
            <http://www.newsreader-project.eu/data/cars/2009/1/19/4VGW-S1T0-TWSV-40J7.xml#
            char=914,920&word=w175&term=t175> .

<http://www.newsreader-project.eu/data/cars/2009/1/19/4VGW-S1T0-TWSV-40J7.xml#pr10,rl21> {
    <http://www.newsreader-project.eu/data/cars/2009/1/19/4VGW-S1T0-TWSV-40J7.xml#workEvent>
            <http://www.newsreader-project.eu/ontologies/propbank/A0>
                    <http://www.newsreader-project.eu/data/cars/2009/1/19/4VGW-S1T0-TWSV-40J7.xml
                    #The+graduate+engineer> .
}

<http://www.newsreader-project.eu/data/cars/2009/1/19/4VGW-S1T0-TWSV-40J7.xml#pr10,rl22> {
    <http://www.newsreader-project.eu/data/cars/2009/1/19/4VGW-S1T0-TWSV-40J7.xml#workEvent>
            <http://www.newsreader-project.eu/ontologies/propbank/A2>
                    <http://www.newsreader-project.eu/data/cars/2009/1/19/4VGW-S1T0-TWSV-40J7.xml
                    #for+VW+group> .
}
```

Figure 7: SRL: Dynamic event with a change of the *work-for* relation as post condition (only FrameNet references)

```
Last week, former Kia managing director Paul Williams,
recruited by SsangYong's importer Koelliker UK last month, hired Cosmic ID to handle its ad account.

  <predicate id="pr27">
      <!--hired-->
      <externalReferences>
        <externalRef reference="hire.01" resource="PropBank"/>
        <externalRef reference="Hiring" resource="FrameNet"/>
        <externalRef reference="contextual" resource="EventType"/>
      </externalReferences>
      <span>
        <target id="t160"/>
      </span>
      <role id="rl59" semRole="A0">
        <!--former Kia managing director Paul Williams , recruited by SsangYong 's importer Koelliker UK last month ,-->
        <externalReferences>
                  <externalRef reference="Hiring#Employer" resource="FrameNet"/>
        </externalReferences>
        <span>
          <target id="t143"/><target id="t144"/><target id="t.mw145"/><target id="t147"/>
          <target head="yes" id="t148"/> ...etc... <target id="t159"/>
        </span>
      </role>
      <role id="rl60" semRole="A1">
        <!--Cosmic ID-->
        <externalReferences>
          <externalRef reference="Hiring#Employee" resource="FrameNet"/>
        </externalReferences>
        <span>
          <target id="t161"/>
          <target head="yes" id="t162"/>
        </span>
      </role>
    </predicate>
```

```
           nwr:data/cars/2009/1/19/4VGW-S1T0-TWSV-40J7.xml#The+graduate+engineer
           nwr:ontology/employedAt
           nwr:data/cars/2009/1/19/4VGW-S1T0-TWSV-40J7.xml#for+VW+group.
       }
```

In the case of the dynamic event *hire*, the following dynamic post-condition rule is triggered (in addition to a pre-condition rule):

1. FrameNet frame *fn:Hiring* is mapped to the dynamic event class JoiningAnOrganization;

2. The roles framenet/Hiring#Employee and /framenet/Hiring#Employer map to nwr:employment-employee and nwr:employment-employer;

3. The post_JoiningAnOrganization_assertion_1 assertion is applied which generates the new triple:

Figure 8: SEM: Dynamic event with a change of the *work-for* relation as post condition

```
<http://www.newsreader-project.eu/data/cars/2008/2/13/4RVV-V1S0-TX4T-010S.xml#hireEvent>
            a               sem:Event , fn:Hiring ;
            rdfs:label      "hire" ;
            gaf:denotedBy  <http://www.newsreader-project.eu/data/cars/2008/2/13/4RVV-V1S0-TX4T-010S.xml#
            char=846,851&word=w160&term=t160> .

    <http://www.newsreader-project.eu/data/cars/2008/2/13/4RVV-V1S0-TX4T-010S.xml#e12>
            a               sem:Actor , <http://www.newsreader-project.eu/ontologies/organization> ;
            rdfs:label      "Cosmic ID" ;
            gaf:denotedBy  <http://www.newsreader-project.eu/data/cars/2008/2/13/4RVV-V1S0-TX4T-010S.xml#
            char=852,861&word=w161,w162&term=t161,t162> .

<http://www.newsreader-project.eu/data/cars/2008/2/13/4RVV-V1S0-TX4T-010S.xml#pr27,rl59> {
    <http://www.newsreader-project.eu/data/cars/2008/2/13/4RVV-V1S0-TX4T-010S.xml#hireEvent>
            <http://www.newsreader-project.eu/ontologies/framenet/Hiring#Employer>
                    <http://dbpedia.org/resource/Paul_Williams_(songwriter)> .
}

<http://www.newsreader-project.eu/data/cars/2008/2/13/4RVV-V1S0-TX4T-010S.xml#pr27,rl60> {
    <http://www.newsreader-project.eu/data/cars/2008/2/13/4RVV-V1S0-TX4T-010S.xml#hireEvent>
            <http://www.newsreader-project.eu/ontologies/framenet/Hiring#Employee>
                    <http://www.newsreader-project.eu/data/cars/2008/2/13/4RVV-V1S0-TX4T-010S.xml#e12> .
}
```

```
nwr:data/cars/inferred-graph2>{
      dbp:resource/Paul_Williams_(songwriter)
      nwr:ontology/employedAt
      nwr:data/cars/2008/2/13/4RVV-V1S0-TX4T-010S.xml#e12.
}
```

The resulting property structure is thus the same across the different expressions and SEM representations. The during and post assertion conditions will however have different implication for the time frame that is inferred for the property. The static event implies that the period associated with the static event also applies as the period for the employedAt property. The dynamic event implies that the time point associated with it marks the beginning of the employedAt property.

The above examples make clear that the mapping of predicates and roles to the ontology is crucial for the recall of the system. This depends on the coverage and precision of the PredicateMatrix and the Semantic Role Labeler. Given the properties that are now defined in the ontology, we can focus the domain adaptation to the processing of these expressions by the NLP modules. In the appendix 10, we provide more examples of expressions of static and dynamic events and the analysis of the current version of the NewsReader pipeline. These examples show that the mapping of the events to the correct properties is not so trivial and considerable work is needed. The main problems are:

1. Lacking mappings to FrameNet frames and elements (see above the example of *own*);
2. Ambiguity of mappings to FrameNet frames and elements; an example of ambiguity is the verb *rise*, which got 7 different frames and is assigned with many different elements that trigger various rules:

```
<externalRef reference="Change_posture" resource="FrameNet"/>
<externalRef reference="Getting_up" resource="FrameNet"/>
<externalRef reference="Motion_directional" resource="FrameNet"/>
<externalRef reference="Path_shape" resource="FrameNet"/>
<externalRef reference="Sidereal_appearance" resource="FrameNet"/>
<externalRef reference="Change_position_on_a_scale" resource="FrameNet"/>
<externalRef reference="Dough_rising" resource="FrameNet"/>
```

3. Complex expressions that combine statements with other information. The appendix 10 shows how they are represented now. In the following examples, the simple property changes are difficult to detect because they are embedded in other complex expressions (marked in bold here):
   - Ford **exercised its right to** buy the Rover trademark under the terms of **its purchase of** Land Rover in 2000, largely as a defensive measure to protect the Land Rover brand.
   - Volkswagen **intends to** increase its market share in Chile to 10% in six years, from the existing 2%.
   - 44-year old Audi (China) Enterprise Management Co., Ltd. General Manager Dr. Dietmar Voggenreiter has been appointed **to take charge of** Chinese business of Audi completely from March 1.
   - **At its eponymous Toyota division**, total sales in February rose 4.4 percent from last year to 149,038 units, and Lexus division total sales grew nearly 4 percent to 17,339 units.
4. Indirect expressions through nominalizations, where the verb is the too general to be mapped to the specific events in the ontology
   - Mr. Foster got a position at GM.
   - Mr. Foster started his job at GM.

7.3 Domain adapted interfacing

Changes in property values provide new ways of interfacing to event data. Rather than taking the events themselves as a starting point, it is now possible to consider simple graphs of value changes over time and/or across regions. Such a graph can be drawn for a single actor or any group of actors (taking the average value of a group). Next, we can place events on the same graph in connection to the value, where we can make a distinction between two types of events:

- Events that directly lead to/imply a value change and the people/organisations involved;
- Events indirectly related to these changes and those involved: causal, temporal, spatial, participatory.

In Figure 9, we show a sketch of a graph that presents values of shares of Volkswagen and Porsche and vice versa in time. Such values can be obtained from structured data or databases but also directly found in the text or indirectly inferred from the text through a reasoner. If the NewsReader program detects static or dynamic events that can be used to infer a property value, these events will be connected as first-order events to the graph. In Figure 9, these first-order events are represented as an inner circle and colored red. Other events that do not directly imply these property values but are connected in some other way to the first-order events can then be grouped in a second circle (colored blue here). Such a grouping of events thus provides a strong ordering of events to be explored by users for potential correlations with properties of the graph, such as extreme values or abrupt changes.

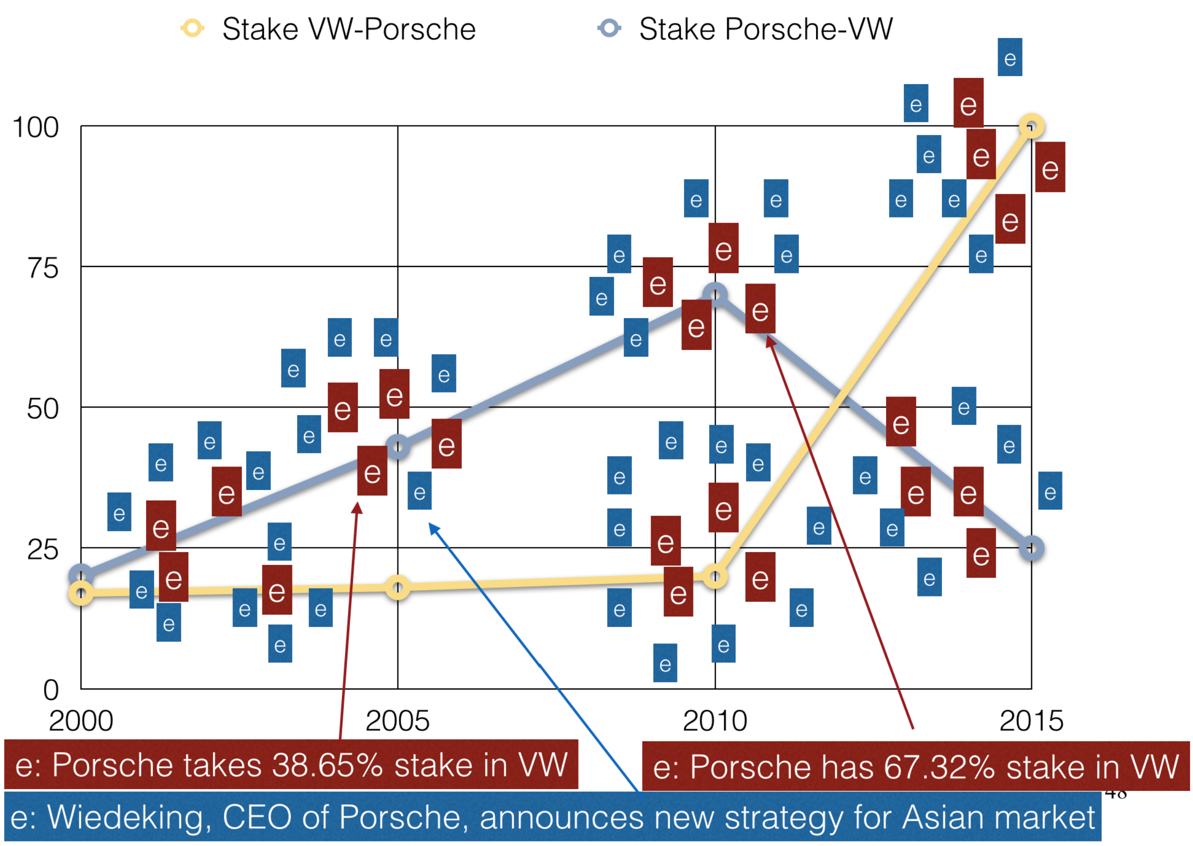

Figure 9: Visualization of value changes over time with directly (red) and indirectly (blue) related events

The types of graphs and properties that can be combined is open to the way the user queries the data. In Figure 10, we see a combination of different properties graphs that can be combined and correlated.

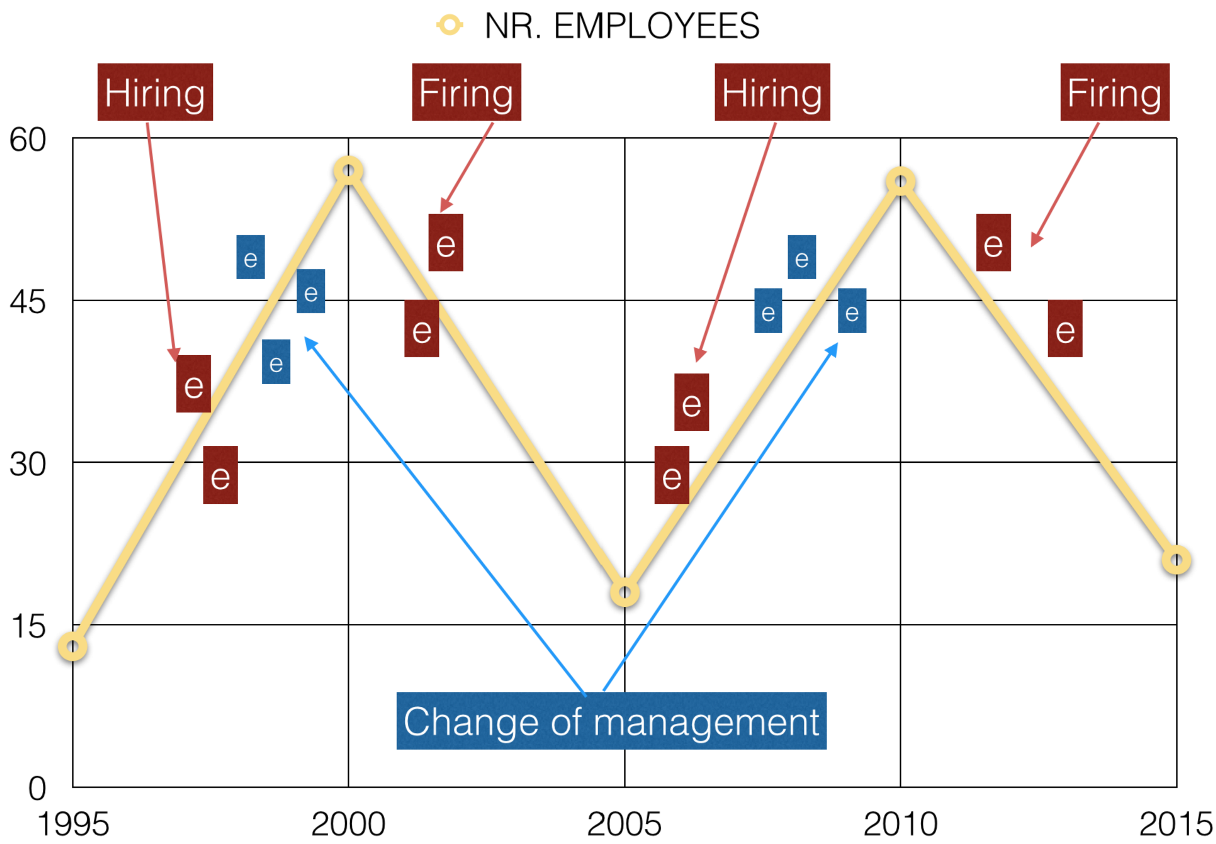

Figure 10: Visualization of value changes over time for different types of properties

The above options for interfacing to value changes will be explored in an upgrade of the Decision Support Tool Suite that is released before the end of the 2nd year of the project.

8 Conclusions and Future Work

In this deliverable we have presented an analysis of the scope and domain mismatches in the first two use cases in the NewsReader dataset. In Section 3 the analysis was focused on entity recognition and linking, from which we concluded that our recognition and linking methods would benefit from the inclusion of additional domain specific knowledge in the form of both resources as well as learning a model from the data itself, as well as extending the entity types.

In Section 4, we present our analyses of the results of the generic English pipeline on events and roles on the global automotive industry dataset. The most important finding here is that the structure that is provided by the resources that events are linked to (such as PropBank and FrameNet) is not sufficient to express the types of relationships between events that we would like to express. To remedy this, we have developed an Event and Situations Ontology that expresses implied changes in the world caused by the events detected in our data (Section 4.3).

Additional vocabularies that may aid the domain adaptation of our pipeline are described in Section 5 and the integration of our findings for entity recognition and linking, the event and situations ontology and the vocabularies with the event recognition and modelling pipeline are described in Section 7.

In the current version of the NLP pipeline (see Deliverable 4.2.2), the Named Entity Recognition module has been extended to include the product category. Integrating additional resources and word representations is currently underway with the reprocessing of the global automotive industry articles data. With the benchmark analysis done in the coming months, the postprocessing steps to filter the final results can be devised.

To obtain a precise overview of the interactions between the modules responsible for relating entities to events and events to each other, we aim to firstly generate a dataset that only includes events described in the Events and Situations Ontology and apply our reasoning techniques over. This will enable us to assess whether the structure of the ESO is finegrained enough and what the interesting avenues for extending this first version of the ontology are.

We are also planning our first domain adaptation experiments to the other project languages. A million documents from a variety of sources concerning the Financial Crisis has been supplied to us by the information specialists at the Dutch House of Representatives. We are investigating co-development of the domain adaptation for English and Dutch to investigate the feasibility of porting our domain adaptation approaches to new languages.

References

Rodrigo Agerri, Itziar Aldabe, Zuhaitz Beloki, Egoitz Laparra, Maddalen Lopez de Lacalle, German Rigau, Aitor Soroa, Marieke van Erp, Piek Vossen, Christian Girardi, and Sara Tonelli. Event detection, version 1. Deliverable D4.2.1, NewsReader, 2013.

Rodrigo Agerri, Josu Bermudez, and German Rigau. IXA Pipeline: Efficient and ready to use multilingual NLP tools. In *Proceedings of the 9th Language Resources and Evaluation Conference (LREC2014), Reykjavik, Iceland*, 2014.

J. Álvez, J. Atserias, J. Carrera, S. Climent, A. Oliver, and G. Rigau. Consistent annotation of eurowordnet with the top concept ontology. In *Proceedings of Fourth International WordNet Conference (GWC'08)*, 2008.

Collin F. Baker, Charles J. Fillmore, and John B. Lowe. The berkeley framenet project. In *Proceedings of the 17th international conference on Computational linguistics (COLING'98)*, volume 1, pages 86–90, 1998.

Luisa Bentivogli, Pamela Forner, Bernardo Magnini, and Emanuele Pianta. Revising the wordnet domains hierarchy: semantics, coverage and balancing. In *Proceedings of the Workshop on Multilingual Linguistic Ressources*, pages 101–108. Association for Computational Linguistics, 2004.

Anders Björkelund, Love Hafdell, and Pierre Nugues. Multilingual semantic role labeling. In *Proceedings of The Thirteenth Conference on Computational Natural Language Learning (CoNLL-2009)*, pages 43–48, Boulder, CO, USA, June 4-5 2009.

Peter F Brown, Peter V Desouza, Robert L Mercer, Vincent J Della Pietra, and Jenifer C Lai. Class-based n-gram models of natural language. *Computational linguistics*, 18(4):467–479, 1992.

Nancy Chinchor and Patricia Robinson. Muc-7 named entity task definition. In *Proceedings of the 7th Conference on Message Understanding*, 1997.

Alexander Clark. Combining distributional and morphological information for part of speech induction. In *Proceedings of the tenth conference on European chapter of the Association for Computational Linguistics-Volume 1*, pages 59–66. Association for Computational Linguistics, 2003.

Maddalen López de Lacalle, Egoitz Laparra, and German Rigau. Predicate matrix: extending semlink through wordnet mappings. In *Proceedings of the 9th international conference on Language Resources and Evaluation (LREC 2014)*, Reykjavik, Iceland, 2014.

Christiane Fellbaum. *WordNet: An Electronic Lexical Database*. Wiley, 1998.

Aitor Gonzalez-Agirre, Egoitz Laparra, and German Rigau. Multilingual central repository version 3.0. In *LREC*, pages 2525–2529, 2012.

Karin Kipper, Hoa Trang Dang, and Martha Palmer. Class-based construction of a verb lexicon. In *AAAI/IAAI*, pages 691–696, 2000.

Adam Meyers, Ruth Reeves, Catherine Macleod, Rachel Szekely, Veronika Zielinska, Brian Young, and Ralph Grishman. The nombank project: An interim report. In *HLT-NAACL 2004 workshop: Frontiers in corpus annotation*, pages 24–31, 2004.

Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*, 2013.

Ian Niles and Adam Pease. Towards a standard upper ontology. In *Proceedings of the international conference on Formal Ontology in Information Systems-Volume 2001*, pages 2–9. ACM, 2001.

Martha Palmer, Dan Gildea, and Paul Kingsbury. The proposition bank: A corpus annotated with semantic roles. *Computational Linguistics Journal*, 31(1), 2005.

Martha Palmer. Semlink: Linking propbank, verbnet and framenet. In *Proceedings of the Generative Lexicon Conference*, pages 9–15, 2009.

Alexandre Passos, Vineet Kumar, and Andrew McCallum. Lexicon infused phrase embeddings for named entity resolution. In *Proceedings of the Conference on Computational Natural Language Learning*, 2014.

L. Ratinov and D. Roth. Design challenges and misconceptions in named entity recognition. In *Proceedings of the Thirteenth Conference on Computational Natural Language Learning*, pages 147–155, 2009.

Josef Ruppenhofer, Michael Ellsworth, Miriam R. L. Petruck, Christopher R. Johnson, and Jan Scheffczyk. *FrameNet II: Extended Theory and Practice.* International Computer Science Institute, Berkeley, California, 2006.

Erik F Tjong Kim Sang and Fien De Meulder. Introduction to the conll-2003 shared task: Language-independent named entity recognition. In *Proceedings of the seventh conference on Natural language learning at HLT-NAACL 2003-Volume 4*, pages 142–147. Association for Computational Linguistics, 2003.

Joseph Turian, Lev Ratinov, and Yoshua Bengio. Word representations: a simple and general method for semi-supervised learning. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pages 384–394. Association for Computational Linguistics, 2010.

Willem Robert van Hage and Thomas Ploeger. System design decision support tools. Deliverable 7.1, NewsReader, 2013.

Piek Vossen, Agata Cybulska, Egoitz Laparra, Oier Lopez de Lacalle, Eneko Agirre, and German Rigau. D5.1.1 event narrative module, version 1. Deliverable, NewsReader, 2013.

9 Appendix: automotive frames

Table 14: Contextual frames occurring in the automotive industry data set

Absorb_heat	Change_position_on_a_scale	Fleeing	Prohibiting
Abundance	Change_posture	Fluidic_motion	Proliferating_in_number
Abusing	Change_tool	Forging	Provide_lodging
Adding_up	Closure	Forming_relationships	Quarreling
Adjusting	Collaboration	Friction	Quitting
Adorning	Colonization	Frugality	Quitting_a_place
Aging	Come_together	Gathering_up	Reading
Amalgamation	Coming_up_with	Getting	Receiving
Amounting_to	Commerce_buy	Getting_up	Recording
Apply_heat	Commerce_collect	Giving	Recovery
Arranging	Commerce_pay	Grinding	Rejuvenation
Arriving	Commerce_sell	Grooming	Releasing
Assemble	Compatibility	Hiding_objects	Removing
Assistance	Competition	Hiring	Render_nonfunctional
Attaching	Compliance	Hit_target	Renting
Attack	Conquering	Holding_off_on	Renting_out
Avoiding	Cooking_creation	Hostile_encounter	Replacing
Becoming_a_member	Corroding	Imitating	Reshaping
Becoming_detached	Corroding_caused	Immobilization	Residence
Behind_the_scenes	Cotheme	Impact	Resolve_problem
Being_attached	Create_physical_artwork	Imprisonment	Resurrection
Being_employed	Create_representation	Inchoative_attaching	Revenge
Being_in_category	Creating	Inchoative_change_of_temperature	Rewards_and_punishments
Being_in_operation	Cure	Ingest_substance	Ride_vehicle
Being_located	Cutting	Ingestion	Robbery
Body_movement	Damaging	Inhibit_movement	Rope_manipulation
Breathing	Daring	Inspecting	Rotting
Bringing	Death	Installing	Scouring
Building	Defend	Institutionalization	Scrutiny
Bungling	Delivery	Intentional_traversing	Seeking
Catastrophe	Departing	Intentionally_create	Self_motion
Cause_change	Destroying	Kidnapping	Sending
Cause_change_of_consistency	Detaching	Killing	Separating
Cause_change_of_phase	Dimension	Knot_creation	Setting_fire
Cause_change_of_position_on_a_scale	Dispersal	Leadership	Shoot_projectiles
Cause_change_of_strength	Dodging	Light_movement	Shopping
Cause_expansion	Dressing	Limiting	Sign_agreement
Cause_fluidic_motion	Duplication	Location_of_light	Similarity
Cause_harm	Earnings_and_losses	Locative_relation	Sleep
Cause_impact	Eclipse	Make_acquaintance	Smuggling
Cause_motion	Education_teaching	Make_noise	Soaking
Cause_temperature_change	Elusive_goal	Manipulate_into_doing	Social_event
Cause_to_amalgamate	Emitting	Manipulation	Sound_movement
Cause_to_be_dry	Employing	Manufacturing	Storing
Cause_to_be_sharp	Emptying	Mass_motion	Supply
Cause_to_be_wet	Escaping	Motion	Surpassing
Cause_to_experience	Evading	Motion_directional	Surviving
Cause_to_fragment	Examination	Motion_noise	Take_place_of
Cause_to_make_noise	Exchange	Moving_in_place	Taking
Cause_to_make_progress	Exchange_currency	Operate_vehicle	Text_creation
Cause_to_move_in_place	Exclude_member	Operational_testing	Theft
Cause_to_start	Excreting	Path_shape	Translating
Cause_to_wake	Execution	Perception	Travel
Change_direction	Expansion	Personal_relationship	Traversing
Change_event_duration	Expensiveness	Piracy	Undergo_change
Change_event_time	Experience_bodily_harm	Placing	Undressing
Change_of_consistency	Experiencer_obj	Posture	Use_firearm
Change_of_leadership	Filling	Precipitation	Visiting
Change_of_phase	Fining	Preserving	Waiting
Change_operational_state	Firing	Processing_materials	Waking_up, Wearing, Weather

Table 15: Source introducing frames

COGNITIVE FRAMES	Seeking	Confronting_problem	Predicting
Adopt_selection	Taking_sides	Contacting	Prevarication
Assessing	Topic	Criminal_investigation	Prohibiting
Awareness	SPEECH_ACT FRAMES	Deny_permission	Questioning
Becoming_aware	Achieving_first	Deserving	Referring_by_name
Categorization	Adding_up	Discussion	Regard
Cause_emotion	Adducing	Distinctiveness	Reporting
Certainty	Agree_or_refuse_to_act	Encoding	Request
Choosing	Appointing	Eventive_cognizer_affecting	Respond_to_proposal
Cogitation	Attempt_suasion	Evidence	Reveal_secret
Coming_to_believe	Bail_decision	Experiencer_obj	Rite
Daring	Be_in_agreement_on_assessment	Expressing_publicly	Seeking
Desiring	Be_translation_equivalent	Forgiveness	Sign
Differentiation	Become_silent	Gesture	Silencing
Emotion_active	Behind_the_scenes	Grant_permission	Simple_naming
Estimating	Being_named	Have_as_translation_equivalent	Speak_on_topic
Expectation	Body_movement	Heralding	Spelling_and_pronouncing
Experiencer_focus	Bragging	Imposing_obligation	Statement
Experiencer_obj	Categorization	Judgment	Suasion
Familiarity	Chatting	Judgment_communication	Subjective_influence
Feeling	Choosing	Judgment_direct_address	Successfully_communicate_message
Feigning	Claim_ownership	Justifying	Talking_into
Grasp	Coming_up_with	Labeling	Telling
Importance	Commitment	Linguistic_meaning	Text_creation
Judgment	Communicate_categorization	Make_agreement_on_action	Verdict
Occupy_rank	Communication	Make_noise	PERCEPTION FRAMES
Opinion	Communication_manner	Making_faces	Appearance
Partiality	Communication_means	Manipulate_into_doing	Categorization
Place_weight_on	Communication_noise	Motion_noise	Chemical-sense_description
Preference	Communication_response	Name_conferral	Locating
Purpose	Compatibility	Notification_of_charges	Perception_active
Reliance	Complaining	Omen	Perception_body
Scrutiny	Compliance	Pardon	Perception_experience

Table 16: Grammatical event frames

Cause_to_end	Inclusion	Relative_time
Coming_to_be	Influence_of_event_on_cognizer	Remainder
Containing	Intentionally_act	Setting_out
Cotheme	Intentionally_affect	Sidereal_appearance
Creating	Launch_process	State_continue
Detaining	Left_to_do	Storing
Emanating	Manipulate_into_doing	Success_or_failure
Event	Operating_a_system	Successful_action
Evidence	Permitting	Taking_time
Execute_plan	Preventing	Taking
Existence	Process_continue	Thriving
Grant_permission	Process_end	Thwarting
Halt	Process_resume	Topic
Have_as_requirement	Process_start	Undergo_change
Hindering	Process_stop	Using
Holding_off_on	Reasoning	

10 Appendix: Examples of property value and property changes expressions

Figure 11: SRL: Static event representing an ownership relation

```
Paris-based Renault SA owns 44\% of Nissan Motor Co.

  <predicate id="pr138">
    <!--owns-->
    <externalReferences>
      <externalRef reference="own.01" resource="PropBank"/>
      <externalRef reference="own-100" resource="VerbNet"/>
      <externalRef reference="contextual" resource="EventType"/>
    </externalReferences>
    <span>
      <target id="t797"/>
    </span>
    <role id="rl311" semRole="A0">
      <!--Paris-based Renault SA-->
      <externalReferences>
        <externalRef reference="own-100#Pivot" resource="VerbNet"/>
      </externalReferences>
      <span>
        <target id="t794"/>
        <target id="t795"/>
        <target head="yes" id="t796"/>
      </span>
    </role>
    <role id="rl312" semRole="A1">
      <!--44 \% of Nissan Motor Co-->
      <externalReferences>
        <externalRef reference="own-100#Theme" resource="VerbNet"/>
      </externalReferences>
      <span>
        <target id="t798"/>
        <target head="yes" id="t799"/>
        <target id="t800"/>
        <target id="t801"/>
        <target id="t802"/>
        <target id="t803"/>
      </span>
    </role>
  </predicate>
```

Figure 12: SEM: Static event representing an ownership relation

```
<http://www.newsreader-project.eu/data/cars/2009/1/14/4VCW-RCH0-TXDB-007C.xml#pr138,rl311> {
    <http://www.newsreader-project.eu/data/cars/2009/1/14/4VCW-RCH0-TXDB-007C.xml#ownEvent>
            <http://www.newsreader-project.eu/ontologies/propbank/A0>
                    <http://www.newsreader-project.eu/data/cars/2009/1/14/4VCW-RCH0-TXDB-007C.xml#
                    Paris-based+Renault+SA> ;
            <http://www.newsreader-project.eu/ontologies/verbnet/own-100#Pivot>
                    <http://www.newsreader-project.eu/data/cars/2009/1/14/4VCW-RCH0-TXDB-007C.xml#
                    Paris-based+Renault+SA> .
}

<http://www.newsreader-project.eu/data/cars/2009/1/14/4VCW-RCH0-TXDB-007C.xml#pr138,rl312> {
    <http://www.newsreader-project.eu/data/cars/2009/1/14/4VCW-RCH0-TXDB-007C.xml#ownEvent>
            <http://www.newsreader-project.eu/ontologies/propbank/A1>
                    <http://www.newsreader-project.eu/data/cars/2009/1/14/4VCW-RCH0-TXDB-007C.xml#
                    44+\%25+of+Nissan+Motor+Co> ;
            <http://www.newsreader-project.eu/ontologies/verbnet/own-100#Theme>
                    <http://www.newsreader-project.eu/data/cars/2009/1/14/4VCW-RCH0-TXDB-007C.xml#
                    44+\%25+of+Nissan+Motor+Co> .
}

<http://www.newsreader-project.eu/data/cars/2009/1/14/4VCW-RCH0-TXDB-007C.xml#pr140,rl314> {
    <http://www.newsreader-project.eu/data/cars/2009/1/14/4VCW-RCH0-TXDB-007C.xml#ownEvent>
            <http://www.newsreader-project.eu/ontologies/propbank/A0>
                    <http://dbpedia.org/resource/Nissan_Motor_Company> ;
            <http://www.newsreader-project.eu/ontologies/verbnet/own-100#Pivot>
                    <http://dbpedia.org/resource/Nissan_Motor_Company> .
}

<http://www.newsreader-project.eu/data/cars/2009/1/14/4VCW-RCH0-TXDB-007C.xml#pr140,rl315> {
    <http://www.newsreader-project.eu/data/cars/2009/1/14/4VCW-RCH0-TXDB-007C.xml#ownEvent>
            <http://www.newsreader-project.eu/ontologies/propbank/A1>
                    <http://www.newsreader-project.eu/data/cars/2009/1/14/4VCW-RCH0-TXDB-007C.xml#
                    15+\%25+of+Renault> ;
            <http://www.newsreader-project.eu/ontologies/verbnet/own-100#Theme>
                    <http://www.newsreader-project.eu/data/cars/2009/1/14/4VCW-RCH0-TXDB-007C.xml#
                    15+\%25+of+Renault> .
}
```

Figure 13: SRL: Static event representing an ownership relation

```
Ford Motor Co . , which owns the Lincoln Mercury brands ,

    <predicate id="pr92">
      <!--owns-->
      <externalReferences>
        <externalRef reference="own.01" resource="PropBank"/>
        <externalRef reference="own-100" resource="VerbNet"/>
        <externalRef reference="contextual" resource="EventType"/>
      </externalReferences>
      <span>
        <target id="t494"/>
      </span>
      <role id="rl199" semRole="A0">
        <!--Ford Motor Co . , which owns the Lincoln Mercury brands ,-->
        <externalReferences>
          <externalRef reference="own-100#Pivot" resource="VerbNet"/>
        </externalReferences>
        <span>
          <target id="t488"/>
          <target head="yes" id="t489"/>
          <target id="t490"/>...ETC... <target id="t499"/>
        </span>
      </role>
      <role id="rl200" semRole="R-A0">
        <!--which-->
        <span>
          <target head="yes" id="t493"/>
        </span>
      </role>
      <role id="rl201" semRole="A1">
        <!--the Lincoln Mercury brands ,-->
        <externalReferences>
          <externalRef reference="own-100#Theme" resource="VerbNet"/>
        </externalReferences>
        <span>
          <target id="t495"/><target id="t496"/><target id="t497"/>
          <target head="yes" id="t498"/>
          <target id="t499"/>
        </span>
      </role>
    </predicate>
```

Figure 14: SEM: Static event representing an ownership relation

```
<http://www.newsreader-project.eu/data/cars/2008/11/9/4TWF-2S00-TXM4-S1KY.xml#pr92,rl199> {
    <http://www.newsreader-project.eu/data/cars/2008/11/9/4TWF-2S00-TXM4-S1KY.xml#ownEvent>
            sem:hasActor  <http://dbpedia.org/resource/Ford_Motor_Company> ;
            <http://www.newsreader-project.eu/ontologies/propbank/A0>
                    <http://dbpedia.org/resource/Ford_Motor_Company> ;
            <http://www.newsreader-project.eu/ontologies/verbnet/own-100#Pivot>
                    <http://dbpedia.org/resource/Ford_Motor_Company> .
}

<http://www.newsreader-project.eu/data/cars/2008/11/9/4TWF-2S00-TXM4-S1KY.xml#pr92,rl201> {
    <http://www.newsreader-project.eu/data/cars/2008/11/9/4TWF-2S00-TXM4-S1KY.xml#ownEvent>
            <http://www.newsreader-project.eu/ontologies/propbank/A1>
                    <http://www.newsreader-project.eu/data/cars/2008/11/9/4TWF-2S00-TXM4-S1KY.xml#the+Lincoln+Mercury+brands+\%2C> ;
            <http://www.newsreader-project.eu/ontologies/verbnet/own-100#Theme>
                    <http://www.newsreader-project.eu/data/cars/2008/11/9/4TWF-2S00-TXM4-S1KY.xml#the+Lincoln+Mercury+brands+\%2C> .
}
```

Figure 15: SRL: Dynamic ownership change

```
Ford exercised its right to buy the Rover trademark under the terms of its purchase of Land Rover in 2000,
largely as a defensive measure to protect the Land Rover brand.

    <predicate id="pr911">
      <!--buy-->
      <externalReferences>
        <externalRef reference="buy.01" resource="PropBank"/>
        <externalRef reference="get-13.5.1" resource="VerbNet"/>
        <externalRef reference="Commerce_buy" resource="FrameNet"/>
        <externalRef reference="contextual" resource="EventType"/>
      </externalReferences>
      <span>
        <target id="t5577"/>
      </span>
      <role id="rl1773" semRole="A0">
        <!--Ford-->
        <externalReferences>
          <externalRef reference="get-13.5.1#Agent" resource="VerbNet"/>
          <externalRef reference="Commerce_buy#Buyer" resource="FrameNet"/>
        </externalReferences>
        <span>
          <target head="yes" id="t5572"/>
        </span>
      </role>
      <role id="rl1774" semRole="A1">
        <!--the Rover trademark-->
        <externalReferences>
          <externalRef reference="get-13.5.1#Theme" resource="VerbNet"/>
          <externalRef reference="Commerce_buy#Goods" resource="FrameNet"/>
        </externalReferences>
        <span>
          <target id="t5578"/>
          <target id="t5579"/>
          <target head="yes" id="t5580"/>
        </span>
      </role>
      <role id="rl1775" semRole="AM-MNR">
        <!--under the terms of its purchase of Land Rover in 2000-->
        <span>
          <target head="yes" id="t5581"/>
          <target id="t5582"/>...ETC... <target id="t5591"/>
        </span>
      </role>
    </predicate>
```

Figure 16: SEM: Dynamic ownership change

```
<http://www.newsreader-project.eu/data/cars/2009/9/21/7WPF-JKT1-2PS2-62DY.xml#pr911,rl1773> {
    <http://www.newsreader-project.eu/data/cars/2009/9/21/7WPF-JKT1-2PS2-62DY.xml#buyEvent>
            <http://www.newsreader-project.eu/ontologies/framenet/Commerce_buy#Buyer>
                    <http://dbpedia.org/resource/Ford_Motor_Company> .
}

<http://www.newsreader-project.eu/data/cars/2009/9/21/7WPF-JKT1-2PS2-62DY.xml#pr911,rl1774> {
    <http://www.newsreader-project.eu/data/cars/2009/9/21/7WPF-JKT1-2PS2-62DY.xml#buyEvent>
            <http://www.newsreader-project.eu/ontologies/framenet/Commerce_buy#Goods>
                    <http://www.newsreader-project.eu/data/cars/2009/9/21/7WPF-JKT1-2PS2-62DY.xml#the+Rover+trademark> ; .
}

<http://www.newsreader-project.eu/data/cars/2009/9/21/7WPF-JKT1-2PS2-62DY.xml#pr1599,rl3151> {
    <http://www.newsreader-project.eu/data/cars/2009/9/21/7WPF-JKT1-2PS2-62DY.xml#buyEvent>
            <http://www.newsreader-project.eu/ontologies/framenet/Commerce_buy#Goods>
                    <http://dbpedia.org/resource/Rover_Company>  .
}
```

Figure 17: SRL: Change of value on a scale for market share

```
Volkswagen intends to increase its market share in Chile to 10\% in six years, from the existing 2\%.

  <predicate id="pr11">
    <!--increase-->
    <externalReferences>
      <externalRef reference="increase.01" resource="PropBank"/>
      <externalRef reference="other_cos-45.4" resource="VerbNet"/>
      <externalRef reference="calibratable_cos-45.6" resource="VerbNet"/>
      <externalRef reference="calibratable_cos-45.6-1" resource="VerbNet"/>
      <externalRef reference="Change_position_on_a_scale" resource="FrameNet"/>
      <externalRef reference="contextual" resource="EventType"/>
    </externalReferences>
    <span>
      <target id="t59"/>
    </span>
    <role id="rl26" semRole="A0">
      <!--Volkswagen-->
      <externalReferences>
        <externalRef reference="other_cos-45.4#Agent" resource="VerbNet"/>
      </externalReferences>
      <span>
        <target head="yes" id="t56"/>
      </span>
    </role>
    <role id="rl27" semRole="A1">
      <!--its market share in Chile-->
      <externalReferences>
        <externalRef reference="other_cos-45.4#Patient" resource="VerbNet"/>
        <externalRef reference="calibratable_cos-45.6#Patient" resource="VerbNet"/>
        <externalRef reference="Change_position_on_a_scale#Attribute" resource="FrameNet"/>
        <externalRef reference="Change_position_on_a_scale#Item" resource="FrameNet"/>
      </externalReferences>
      <span>
        <target id="t60"/><target id="t61"/>
        <target head="yes" id="t62"/>
        <target id="t63"/><target id="t64"/>
      </span>
    </role>
    <role id="rl28" semRole="A4">
      <!--to 10 \%-->
      <span>
        <target head="yes" id="t65"/><target id="t66"/><target id="t67"/>
      </span>
    </role>
    <role id="rl29" semRole="AM-TMP">
      <!--in six years-->
      <span>
        <target head="yes" id="t68"/><target id="t69"/><target id="t70"/>
      </span>
    </role>
    <role id="rl30" semRole="A3">
      <!--from the existing 2 %-->
```

Figure 18: SEM: Change of value on a scale for market share

```
<http://www.newsreader-project.eu/data/worldcup/2013/2/28/57VV-5311-F111-G0HJ.xml#pr11,rl27> {
    <http://www.newsreader-project.eu/data/worldcup/2013/2/28/57VV-5311-F111-G0HJ.xml#increaseEvent>
            <http://www.newsreader-project.eu/ontologies/framenet/Change_position_on_a_scale#Attribute>
                    <http://www.newsreader-project.eu/data/worldcup/2013/2/28/57VV-5311-F111-G0HJ.xml#its+market+share+in+Chile> ;
            <http://www.newsreader-project.eu/ontologies/framenet/Change_position_on_a_scale#Item>
                    <http://www.newsreader-project.eu/data/worldcup/2013/2/28/57VV-5311-F111-G0HJ.xml#its+market+share+in+Chile>  .
}

<http://www.newsreader-project.eu/data/worldcup/2013/2/28/57VV-5311-F111-G0HJ.xml#pr11,rl28> {
    <http://www.newsreader-project.eu/data/worldcup/2013/2/28/57VV-5311-F111-G0HJ.xml#increaseEvent>
            <http://www.newsreader-project.eu/ontologies/propbank/A4>
                    <http://www.newsreader-project.eu/data/worldcup/2013/2/28/57VV-5311-F111-G0HJ.xml#to+10+\%25> .
}

<http://www.newsreader-project.eu/data/worldcup/2013/2/28/57VV-5311-F111-G0HJ.xml#pr11,rl30> {
    <http://www.newsreader-project.eu/data/worldcup/2013/2/28/57VV-5311-F111-G0HJ.xml#increaseEvent>
            <http://www.newsreader-project.eu/ontologies/propbank/A3>
                    <http://www.newsreader-project.eu/data/worldcup/2013/2/28/57VV-5311-F111-G0HJ.xml#from+the+existing+2+\%25> .
}
```

Figure 19: default

Figure 20: SRL: Change of business position

```
44-year old Audi (China) Enterprise Management Co., Ltd. General Manager Dr. Dietmar Voggenreiter
has been appointed to take charge of Chinese business of Audi completely from March 1.

    <predicate id="pr1">
      <!--appointed-->
      <externalReferences>
        <externalRef reference="appoint.01" resource="PropBank"/>
        <externalRef reference="appoint-29.1" resource="VerbNet"/>
        <externalRef reference="Change_of_leadership" resource="FrameNet"/>
        <externalRef reference="contextual" resource="EventType"/>
      </externalReferences>
      <span>
        <target id="t31"/>
      </span>
      <role id="rl1" semRole="A1">
        <!--General Manager Dr. Dietmar Voggenreiter-->
        <externalReferences>
          <externalRef reference="appoint-29.1#Theme" resource="VerbNet"/>
          <externalRef reference="Change_of_leadership#New_leader" resource="FrameNet"/>
        </externalReferences>
        <span>
          <target id="t24"/>
          <target id="t25"/>
          <target id="t26"/>
          <target id="t27"/>
          <target head="yes" id="t28"/>
        </span>
      </role>
      <role id="rl2" semRole="A2">
        <!--to take charge of Chinese business of Audi completely from March 1-->
        <externalReferences>
          <externalRef reference="appoint-29.1#Result" resource="VerbNet"/>
          <externalRef reference="Change_of_leadership#Role" resource="FrameNet"/>
        </externalReferences>
        <span>
          <target head="yes" id="t32"/>
          <target id="t.mw33"/>...etc... <target id="t43"/>
        </span>
      </role>
    </predicate>
```

Figure 21: SEM: Change of business position

```
<http://www.newsreader-project.eu/data/cars/2013/2/28/57VM-91J1-JC8H-C3PW.xml#appointEvent>
            a               sem:Event , fn:Change_of_leadership ;
            rdfs:label      "appoint" ;
            gaf:denotedBy  <http://www.newsreader-project.eu/data/cars/2013/2/28/57VM-91J1-JC8H-C3PW.xml#
            char=165,174&word=w31&term=t31> .

<http://www.newsreader-project.eu/data/cars/2013/2/28/57VM-91J1-JC8H-C3PW.xml#e7>
            a               sem:Actor , <http://www.newsreader-project.eu/ontologies/person> ;
            rdfs:label      "Manager Dr. Dietmar Voggenreiter" ;
            gaf:denotedBy  <http://www.newsreader-project.eu/data/cars/2013/2/28/57VM-91J1-JC8H-C3PW.xml#
            char=123,155&word=w25,w26,w27,w28&term=t25,t26,t27,t28> .

<http://www.newsreader-project.eu/data/cars/2013/2/28/57VM-91J1-JC8H-C3PW.xml#pr1,rl1> {
    <http://www.newsreader-project.eu/data/cars/2013/2/28/57VM-91J1-JC8H-C3PW.xml#appointEvent>
            <http://www.newsreader-project.eu/ontologies/framenet/Change_of_leadership#New_leader>
                    <http://www.newsreader-project.eu/data/cars/2013/2/28/57VM-91J1-JC8H-C3PW.xml#e7> .
}

<http://www.newsreader-project.eu/data/cars/2013/2/28/57VM-91J1-JC8H-C3PW.xml#pr1,rl2> {
    <http://www.newsreader-project.eu/data/cars/2013/2/28/57VM-91J1-JC8H-C3PW.xml#appointEvent>
                <http://www.newsreader-project.eu/ontologies/framenet/Change_of_leadership#Role>
                    <http://dbpedia.org/resource/China>  .
}
```

Figure 22: SRL: Change of position on a scale for ”sales rose” (only FrameNet)

```
At its eponymous Toyota division, total sales in February rose 4.4 percent from last year to 149,038 units,
and Lexus division total sales grew nearly 4 percent to 17,339 units.
```


```
   <predicate id="pr52">
      <!--rose-->
      <externalReferences>
        <externalRef reference="Change_posture" resource="FrameNet"/>
        <externalRef reference="Getting_up" resource="FrameNet"/>
        <externalRef reference="Motion_directional" resource="FrameNet"/>
        <externalRef reference="Path_shape" resource="FrameNet"/>
        <externalRef reference="Sidereal_appearance" resource="FrameNet"/>
        <externalRef reference="Change_position_on_a_scale" resource="FrameNet"/>
        <externalRef reference="Dough_rising" resource="FrameNet"/>
        <externalRef reference="contextual" resource="EventType"/>
      </externalReferences>
      <span>
        <target id="t259"/>
      </span>
      <role id="rl96" semRole="AM-LOC">
        <!--At its eponymous Toyota division , etc.. .-->
        <span>
          <target head="yes" id="t249"/>
          <target id="t250"/> ....etc... <target id="t281"/>
        </span>
      </role>
      <role id="rl97" semRole="A1">
        <!--total sales in February-->
        <externalReferences>
          <externalRef reference="Change_posture#Protagonist" resource="FrameNet"/>
          <externalRef reference="Motion_directional#Theme" resource="FrameNet"/>
          <externalRef reference="Path_shape#Road" resource="FrameNet"/>
          <externalRef reference="Change_position_on_a_scale#Item" resource="FrameNet"/>
        </externalReferences>
        <span>
          <target id="t255"/>
          <target head="yes" id="t256"/>
          <target id="t257"/>
          <target id="t258"/>
        </span>
      </role>
      <role id="rl98" semRole="A2">
        <!--4.4 percent-->
        <externalReferences>
          <externalRef reference="Change_position_on_a_scale#Difference" resource="FrameNet"/>
        </externalReferences>
        <span>
          <target id="t260"/>
          <target head="yes" id="t261"/>
        </span>
      </role>
      <role id="rl99" semRole="A3">
        <!--from last year-->
```

Figure 23: SEM: Change of position on a scale for "sales rose" (only FrameNet)

```
<http://www.newsreader-project.eu/data/cars/2013/3/1/57W0-0GP1-F0R9-G0BB.xml#pr52,rl96> {
    <http://www.newsreader-project.eu/data/cars/2013/3/1/57W0-0GP1-F0R9-G0BB.xml#riseEvent>
            <http://www.newsreader-project.eu/ontologies/propbank/AM-LOC>
                    <http://dbpedia.org/resource/Toyota> .
}

<http://www.newsreader-project.eu/data/cars/2013/3/1/57W0-0GP1-F0R9-G0BB.xml#pr52,rl97> {
    <http://www.newsreader-project.eu/data/cars/2013/3/1/57W0-0GP1-F0R9-G0BB.xml#riseEvent>
            <http://www.newsreader-project.eu/ontologies/framenet/Change_position_on_a_scale#Item>
                    <http://www.newsreader-project.eu/data/cars/2013/3/1/57W0-0GP1-F0R9-G0BB.xml#
                    total+sales+in+February> ;
            <http://www.newsreader-project.eu/ontologies/framenet/Change_posture#Protagonist>
                    <http://www.newsreader-project.eu/data/cars/2013/3/1/57W0-0GP1-F0R9-G0BB.xml#
                    total+sales+in+February> ;
            <http://www.newsreader-project.eu/ontologies/framenet/Motion_directional#Theme>
                    <http://www.newsreader-project.eu/data/cars/2013/3/1/57W0-0GP1-F0R9-G0BB.xml#
                    total+sales+in+February> ;
            <http://www.newsreader-project.eu/ontologies/framenet/Path_shape#Road>
                    <http://www.newsreader-project.eu/data/cars/2013/3/1/57W0-0GP1-F0R9-G0BB.xml#
                    total+sales+in+February>  .
}

<http://www.newsreader-project.eu/data/cars/2013/3/1/57W0-0GP1-F0R9-G0BB.xml#pr52,rl98> {
    <http://www.newsreader-project.eu/data/cars/2013/3/1/57W0-0GP1-F0R9-G0BB.xml#riseEvent>
            <http://www.newsreader-project.eu/ontologies/framenet/Change_position_on_a_scale#Difference>
                    <http://www.newsreader-project.eu/data/cars/2013/3/1/57W0-0GP1-F0R9-G0BB.xml#
                    4.4+percent> .
}

<http://www.newsreader-project.eu/data/cars/2013/3/1/57W0-0GP1-F0R9-G0BB.xml#pr52,rl99> {
    <http://www.newsreader-project.eu/data/cars/2013/3/1/57W0-0GP1-F0R9-G0BB.xml#riseEvent>
            <http://www.newsreader-project.eu/ontologies/framenet/Motion_directional#Source>
                    <http://www.newsreader-project.eu/data/cars/2013/3/1/57W0-0GP1-F0R9-G0BB.xml#
                    from+last+year> ;
            <http://www.newsreader-project.eu/ontologies/framenet/Path_shape#Source>
                    <http://www.newsreader-project.eu/data/cars/2013/3/1/57W0-0GP1-F0R9-G0BB.xml#
                    from+last+year> .
}

<http://www.newsreader-project.eu/data/cars/2013/3/1/57W0-0GP1-F0R9-G0BB.xml#pr52,rl100> {
    <http://www.newsreader-project.eu/data/cars/2013/3/1/57W0-0GP1-F0R9-G0BB.xml#riseEvent>
            <http://www.newsreader-project.eu/ontologies/framenet/Motion_directional#Goal>
                    <http://www.newsreader-project.eu/data/cars/2013/3/1/57W0-0GP1-F0R9-G0BB.xml#
                    to+149\%2C038+units> ;
            <http://www.newsreader-project.eu/ontologies/framenet/Path_shape#Goal>
                    <http://www.newsreader-project.eu/data/cars/2013/3/1/57W0-0GP1-F0R9-G0BB.xml#
                    to+149\%2C038+units>  .
}
```

Figure 24: SRL: Change position on a scale for "sales grew"

```
At its eponymous Toyota division, total sales in February rose 4.4 percent from last year to 149,038 units,
and Lexus division total sales grew nearly 4 percent to 17,339 units.

    <predicate id="pr53">
      <!--grew-->
      <externalReferences>
        <externalRef reference="Cause_expansion" resource="FrameNet"/>
        <externalRef reference="Change_position_on_a_scale" resource="FrameNet"/>
        <externalRef reference="Expansion" resource="FrameNet"/>
        <externalRef reference="contextual" resource="EventType"/>
      </externalReferences>
      <span>
        <target id="t274"/>
      </span>
      <role id="rl101" semRole="AM-LOC">
        <!--At its eponymous Toyota division , etc... .-->
        <span>
          <target head="yes" id="t249"/>
          <target id="t250"/>... etc...<target id="t281"/>
        </span>
      </role>
      <role id="rl102" semRole="A1">
        <!--Lexus division total sales-->
        <externalReferences>
          <externalRef reference="Cause_expansion#Item" resource="FrameNet"/>
          <externalRef reference="Change_position_on_a_scale#Attribute" resource="FrameNet"/>
          <externalRef reference="Expansion#Item" resource="FrameNet"/>
        </externalReferences>
        <span>
          <target id="t270"/>
          <target id="t271"/>
          <target id="t272"/>
          <target head="yes" id="t273"/>
        </span>
      </role>
      <role id="rl103" semRole="A2">
        <!--nearly 4 percent-->
        <span>
          <target id="t275"/>
          <target id="t276"/>
          <target head="yes" id="t277"/>
        </span>
      </role>
      <role id="rl104" semRole="A4">
        <!--to 17,339 units-->
        <span>
          <target head="yes" id="t278"/>
          <target id="t279"/>
          <target id="t280"/>
        </span>
      </role>
    </predicate>
```

Figure 25: SEM: Change position on a scale for "sales grew"

```
<http://www.newsreader-project.eu/data/cars/2013/3/1/57W0-0GP1-F0R9-G0BB.xml#pr53,rl101> {
    <http://www.newsreader-project.eu/data/cars/2013/3/1/57W0-0GP1-F0R9-G0BB.xml#growEvent>
            <http://www.newsreader-project.eu/ontologies/propbank/AM-LOC>
                    <http://dbpedia.org/resource/Toyota> .
}

<http://www.newsreader-project.eu/data/cars/2013/3/1/57W0-0GP1-F0R9-G0BB.xml#pr53,rl102> {
    <http://www.newsreader-project.eu/data/cars/2013/3/1/57W0-0GP1-F0R9-G0BB.xml#growEvent>
            <http://www.newsreader-project.eu/ontologies/framenet/Cause_expansion#Item>
                    <http://www.newsreader-project.eu/data/cars/2013/3/1/57W0-0GP1-F0R9-G0BB.xml#
                    Lexus+division+total+sales> ;
            <http://www.newsreader-project.eu/ontologies/framenet/Change_position_on_a_scale#Attribute>
                    <http://www.newsreader-project.eu/data/cars/2013/3/1/57W0-0GP1-F0R9-G0BB.xml#
                    Lexus+division+total+sales>  .
}

<http://www.newsreader-project.eu/data/cars/2013/3/1/57W0-0GP1-F0R9-G0BB.xml#pr53,rl103> {
    <http://www.newsreader-project.eu/data/cars/2013/3/1/57W0-0GP1-F0R9-G0BB.xml#growEvent>
            <http://www.newsreader-project.eu/ontologies/propbank/A2>
                    <http://www.newsreader-project.eu/data/cars/2013/3/1/57W0-0GP1-F0R9-G0BB.xml#
                    nearly+4+percent> .
}

<http://www.newsreader-project.eu/data/cars/2013/3/1/57W0-0GP1-F0R9-G0BB.xml#pr53,rl104> {
    <http://www.newsreader-project.eu/data/cars/2013/3/1/57W0-0GP1-F0R9-G0BB.xml#growEvent>
            <http://www.newsreader-project.eu/ontologies/propbank/A4>
                    <http://www.newsreader-project.eu/data/cars/2013/3/1/57W0-0GP1-F0R9-G0BB.xml#
                    to+17\%2C339+units> .
}
```