Cross-lingual event-mining using wordnet as a shared knowledge interface

Piek Vossen
VU University Amsterdam
p.t.j.m.vossen@vu.nl

Aitor Soroa, Beñat Zapirain, German Rigau
University of the Basque Country
a.soroa@ehu.es
benat.zapirain@ehu.es
german.rigau@ehu.es

Abstract

We describe a concept-based event-mining system that maximizes information extracted from text and is not restricted to predefined knowledge templates. Such a system needs to handle a wide range of expressions while being able to extract precise semantic relations. The system uses simple patterns of linguistic and ontological constraints that are applied to a uniform representation of the text. It uses a generic ontology based on DOLCE and wordnets in different languages to extract events from text in these languages in an interoperable way. The system performs unsupervised domain-independent event-mining with promising results. Error-analysis showed that the semantic model and the mapping of text to concepts through word-sense-disambiguation (WSD) are not the main cause of the errors but the complexity of the grammatical structures and the quality of parsing. Using the same semantic model and their cross-wordnet links, our English event-mining patterns were transferred to Dutch in less than a day's work. The platform was tested on the environment domain but can be applied to any other domain.

1 Introduction

Traditionally, Information Extraction (IE) is the task of filling template information from previously unseen text which belongs to a predefined domain (Peshkin and Pfeffer, 2003). Standard IE systems are based on language-specific pattern matching (Kaiser and Miksch, 2005), consisting of language specific regular expressions and associated mappings from syntactic to logical forms. The major disadvantage of traditional IE systems is that they focus on satisfying precise, narrow, pre-specified requests e.g. all names of directors of movies. Compared to full text indexes in Information Retrieval (IR), IE systems only cover a small portion of the knowledge in texts while capturing deeper semantic relations.

Alternatively, the KYOTO system[1] combines the comprehensiveness of IR systems with the depths of IE systems. Furthermore, the system can be applied to different languages and domains in an uniform way. It tries to extract any event and its participants from the text, and relate them to time and place. To achieve this, it uses a full text representation format and a wide range of knowledge in the form of wordnets and a generic ontology (Vossen and Rigau, 2010). Word-sense-disambiguation (WSD) is a crucial step to map text to concepts. We implemented a two-phased WSD strategy and show the effects on event-extraction. We first apply a state-of-the-art WSD to words in context, scoring all possible synsets in a wordnet. Each of these synsets is mapped to a shared ontology. From this mapping, all possible ontological implications are derived. Next, the mining system extracts all possible interpretations of all sequences of ontological concepts that match the patterns. In the second-phase, we select an interpretation only if there is a choice using the WSD score. The system has been tested on texts from the environment domain. However, the knowledge resources and patterns are generic and can be applied to any other domain.

In the next section, we describe the general architecture of the KYOTO system and the knowledge structure. In section 3, we describe the module for mining knowledge from the text. In section 4, we describe the evaluation results and an error-analysis for English. Since the profiles use

[1]Available at http://www.kyoto-project.eu/

```
<kaf xml:lang="en" doc="example1">
 <text>
<wf page="1" sent="40" wid="w267"  fileoffset="6,11">
       water</wf>
<wf page="1" sent="40" wid="w268"  fileoffset="12,21">
      pollution</wf>
  <...>
 </text>
 <terms>
<term id="t241" lemma="water" pos="N">
   <span><target id="w267"></target></span>
   <externalReferences>
     <externalRef conf="0.29" ref="14845743-n"
       resource="wn30g"/>
     <externalRef ref="Kyoto#water"
       reftype="sc_equivalenceOf" resource="ontology"/>
     <!--...-->
     <externalRef reftype="SubClassOf"
       reference="DOLCE-Lite.owl#endurant"
       status="implied"/>
     <!--...-->
  </externalReferences>
</term>
<term  tid="t242" lemma="pollution" pos="N" type="open">
   <span><target id="w268"></target></span>
   <externalReferences>
      <externalRef conf="0.33" ref="14516743-n"
        resource="wn30g"/>
      <externalRef ref="Kyoto#contamination_pollution"
        reftype="sc_equivalenceOf" resource="ontology"/>
      <!--...-->
      <externalRef reftype="SubClassOf"
       reference="DOLCE-Lite.owl#perdurant"
       status="implied"/>
      <!--...-->
  </externalReferences>
</term>
 </terms>
<!-- Additional layers (chunking, dependencies, ...) -->
</kaf>
```

Figure 1: Example of a KAF document

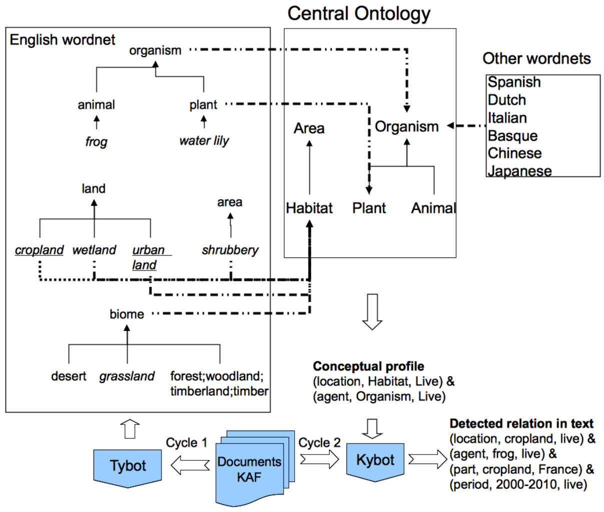

Figure 2: System Architecture

language-neutral ontological constraints, they can be easily transferred to another language. Therefore in section 5, we describe how the system was transferred from English to Dutch, through the wordnet-equivalence links.

2 KYOTO overview

The KYOTO system starts with linguistic processors that apply tokenization, segmentation, morpho-syntactic analysis and semantic tagging of the text. The semantic tagging involves detection of named-entities and the meaning of words according to a given wordnet. The output of the linguistic processors is stored in an XML annotation format that is the same for all the languages, called the KYOTO Annotation Format (KAF, (Bosma et al., 2009)). KAF is compatible with the Linguistic Annotation Framework (LAF, (Ide and L.Romary, 2003)). In KAF, words, terms, constituents and syntactic dependencies are stored in separate layers with references across the structures. All modules in KYOTO draw their input from these XML structures. Likewise, WSD is done on the same KAF annotation in different languages and is therefore the same module for all the languages (Agirre and Soroa, 2009). The current system includes processors for English, Dutch, Italian, Spanish, Basque, Chinese and Japanese. Figure 1 shows a simplified example of a KAF structure with the two basic layers: `<text>` layer (tokenization, segmentation) and `<terms>`, containing morpho-syntactic and semantic information drawn from WordNet and the KYOTO ontology for the words *water* and *pollution*.

In KYOTO, the knowledge extraction is done by so-called Kybots (Knowledge Yielding Robots). Kybots are defined by a set of profiles representing information patterns. In the profile, conceptual relations are expressed using ontological and morpho-syntactic patterns. Since the semantics is defined through the ontology, it is possible to detect similar data even if expressed differently. In Figure 2, we show an example of a conceptual pattern for the environment domain that relates organisms that live in habitats. The pattern uses labels from the central ontology, whereas each wordnet synsets is directly or indirectly related to these labels. Such a pattern can be used to extract events from text, such as *frogs that live in cropland in France during the period 2000-2010*.

The system exploits a 3-layered knowledge-architecture (Vossen and Rigau, 2010), using a central ontology, wordnets in different languages and potential background vocabularies linked to the wordnets. The ontology consists of around 2,000 classes divided over three layers (Hicks and Herold, 2009). The top layer is based on DOLCE[2]

[2]DOLCE-Lite-Plus version 3.9.7

(Gangemi et al., 2003) and OntoWordNet. The second layer are the Base Concepts[3] (BCs) which cover an intermediate level of abstraction for all nominal and verbal WordNet synsets (Izquierdo et al., 2007). Examples of BCs are: *building, vehicle, animal, plant, change, move, size, weight*. A third layer consists of domain classes introduced for detecting events and qualities in a particular domain (i.e. environment).

The semantic model also provides complete mappings to the ontology for all nominal, verbal and adjectival WordNet3.0 synsets (Fellbaum, 1998)[4]. The mappings also harmonize predicate information across different part-of-speech (POS). For instance, migratory events represented by different synsets of the verb *migrate*, the noun *migration* or the adjective *migratory* inherit the same ontological information corresponding to the *ChangeOfResidence* class in the ontology.

This generic knowledge model provides an extremely powerful basis for semantic processing in any domain. Furthermore, through the equivalence relations of wordnets in other languages to the English WordNet, this semantic framework can also be applied to other languages as shown in Section 5.

The WSD module assigns concepts to each word with a score based on the context. Ontological tagging of the text is then the last step in the pre-processing before the extraction of events. For each synset associated to a word, we use the wordnet to ontology mappings to look up its associated ontological classes and inherited properties. The Base Concept mapping guarantees that every synset is mapped to an ontological class. Next, we insert into KAF all the ontological implications that apply to each concept. By making the implicit ontological statements explicit, Kybots are able to find the same relations hidden in different expressions with different surface realizations, e.g.: *water pollution, polluted water, pollution of water, water that is polluted* directly or indirectly express the same relations. Figure 1 shows how ontological statements are represented in KAF as external references related to synsets with a score from the WSD (the value of the atribute *conf*). Words in the term structure usually get many ontological implications for each word meaning. The implications reflect subclass relations from the ontology but also other relations such as events in which the concept denoted by the word plays a role (e.g. the word *polluted water* denotes *water* that plays a role in the event *pollution*) or, the other way around, the roles involved in the event denoted by the word (e.g. the word *water pollution* denotes events in which *water* is as a patient).

[3] `http://adimen.si.ehu.es/web/BLC`

[4] This knowledge model is freely available through the KYOTO website as open-source data.

3 Event extraction

A set of abstract patterns called Kybots use the central ontology to extract actual concept instances and relations from KAF documents. Event-mining is done by processing these abstract patterns on the enriched documents. These patterns are defined in a declarative format using profiles, which describe general morpho-syntactic and semantic conditions on sequences of KAF terms (which are lemmas in the text). These profiles are compiled to XQueries to efficiently scan over KAF documents uploaded into an XML database. These patterns extract the relevant information from each match.

Figure 3 shows an example of a simple Kybot profile. Profiles are described using XML syntax and consist of three main parts:

- Variable declaration (`<variables>` element). In this part, the search entities are defined, e.g.: **X** (terms whose part-of-speech is noun and whose lemma is not “system”), **Y** (terms whose lemma is either “release”, “produce” or “generate”) and **Z** (terms linked to a subclass of the ontological class *DOLCE-Lite.owl#contamination_pollution*, meaning *being contaminated with harmful substances*).
- Relations among variables (`<rel>` element): This part specifies the relations among the previously defined variables, e.g.: **Y** is the main pivot, variable **X** must precede variable **Y** in the same sentence, and variable **Z** must follow variable **Y**. Thus, this relation declares patterns like '**X** → **Y** → **Z**' in a sentence.
- Output template: describes the output to be produced for every match, e.g.: each match generates a new event from the term **Y** and two roles: the 'done-by' role filled by term **X** and 'patient' role, filled by **Z**.

We created 261 generic profiles for English. These profiles capture very simple sequences of parts-of-speech or words, e.g. noun-verb or

```
<kprofile>
 <variables>
  <var name="x" type="term" pos="N" lemma="! system"/>
  <var name="y" type="term"
       lemma="produce | generate | release"/>
  <var name="z" type="term"
       ref="DOLCE-Lite.owl#contamination_pollution"
       reftype="SubClassOf"/>
 </variables>
 <relations>
  <root span="y"/>
  <rel span="x" pivot="y" direction="preceding"/>
  <rel span="z" pivot="y" direction="following"/>
 </relations>
 <events>
  <event target="$y/@tid" lemma="$y/@lemma"
        pos="$y/@pos"/>
  <role target="$x/@tid" rtype="done-by"
        lemma="$x/@lemma"/>
  <role target="$z/@tid" rtype="patient"
        lemma="$z/@lemma"/>
 </events>
</kprofile>
```

Figure 3: Example of a Kybot profile

adjective-noun, where each word is restricted to classes from the ontology, e.g. a motion event followed by a geographical region. Note that it is important that all possible expressions of relation are modeled by the profiles.

4 Evaluation

4.1 Triplet representation for representing events

The event structure in KYOTO is rather specific and events can be complex, including many different roles and relations. Below is an example of such a structure extracted from the sentence: “Forests also absorb air pollution and retain up to 85 percent of the nitrogen from sources such as automobiles and power plants.”

```
<event eid="e203" target="t4260"
   lemma="absorb" pos="V"
   synset="eng-30-01539633-v" rank="0.25"/>
<role rid="r280" event="e203" target="t4258"
   lemma="forest" pos="N" rtype="done-by"
   synset="eng-30-09284015-n" rank="0.15"/>
<role rid="r976" event="e203" target="t4262mw"
   lemma="air pollution" pos="N" rtype="patient"
   synset="eng-30-14517412-n" rank="1"/>
<role rid="r1609" event="e203" target="t4277mw"
   lemma="power plant" pos="N" rtype="simple-cause-of"
   synset="eng-30-03996655-n" rank="1"/>
<role rid="r276" event="e203" target="t4274"
   lemma="automobile" pos="N" rtype="simple-cause-of"
   synset="eng-30-02958343-n" rank="1"/>
```

To be able to compare our results with the output of other systems and gold-standards, we defined a more neutral and simple triplet format.

A triplet consists of:

- a relation
- a list of text token ids that represent the event
- a list of text token ids that represent a participant

If an event has multiple participants, a separate triplet is created for each event-participant pair. The triplet identifier is used to mark which triplets relate to the same event.

4.2 Evaluation results for English

We created a gold-standard in the triplet format. An annotation tool that reads KAF and can assign any set of tags to tokens in KAF was used to make a gold-standard for a document about the Chesapeake Bay, a large estuary in the US[5]. The document has $16,145$ word tokens. We manually annotated all relevant relations in 127 sentences, corresponding to $1,416$ tokens, 353 triplets and 201 events.

The first column in Table 2 show the annotated relations.[6] The patient relation is most frequent (38%), followed by done-by (15%) and simple-cause-of (14%).

As a baseline, we created triplets for all heads of constituents in a single sentence according to the constituent representation of the text in KAF. The baseline generates $3,427$ triplets for the annotated sentences. Since there is no relation predicted, we assume the most-frequent patient relation.

To evaluate the Kybots, we used the 261 generic profiles. The profiles generated 548 triplets for the annotated sentences. In total 169 profiles or combinations of profiles (since multiple profiles can propose the same triplet) have been applied to the annotated fragment.

To measure the proportion of relevant events that are detected by these heuristics, we compare the baseline and Kybot events with the event tokens in the gold standard. The gold standard has 201 events and the baseline 1,627 events, of which 249 overlap with the gold standard events. This results in a recall of 1.24 and a precision of 0.15. The Kybot profiles detect 733 events, of which 209 are relevant. Recall is 1.04 and precision is 0.29. Recall of events is similar to the baseline and precision is almost twice as high. The fact that the recall is higher than 1 is caused by the fact that the gold standard sometimes marks larger phrases as a single event which may be separate events in

[5]The tool and evaluation data is available at the KYOTO website

[6]The relations are taken from the DOLCE part of the KYOTO ontology. Please consult the DOLCE ontology for their formal definition.

the baseline and the Kybot output. Both the baseline and the Kybot profiles thus do not miss any relevant events but do extract a substantial amount of irrelevant events. The precision for the profiles is still reasonable, given the fact that no relevance ranking has been applied and only generic profiles have been used. Note that events that are not annotated can still be proper events.

Table 1 shows the results for the triplet evaluation of the relevant events.

	Ignoring relations		With relations	
	Baseline	Kybots	Baseline	Kybots
Nr. correct	306	222	115	174
Precision	0.09	0.49	0.03	0.32
Recall	0.86	0.63	0.33	0.49

Table 1: Baseline and Kybot results

When ignoring the relation, recall for the baseline is 86%, which shows that the baseline matches a substantial part of the annotated triplets. It also shows that 14% is missed. This is due to the fact that the parser only marks one word as the head in the case of a coordination of heads, e.g. in the phrase "birds and fish" only "bird" is marked as the head. Precision of the baseline is very low, even when we ignore the relation itself. If we take the patient relation as the default, we see that the precision and recall drop even more. The Kybot profiles clearly outperform the baseline in terms of precision: 49% when ignoring the relation and 32% considering all relations. In terms of recall, we see that 63% is covered when we ignore the relation. This indicates that the profiles do consider the majority of structures, but still miss 37% of the structures. When we consider the relations, recall drops to 49% which is still well above the baseline.

4.2.1 Error analysis

We did a separate error analysis for recall and precision. First of all, we checked the 1,023 term tokens of content words (nouns, verbs and adjectives) that occurred in the 127 gold-standard sentences. It turned out that there are 70 tokens with the wrong POS assigned (7%). The major errors are nouns and verbs interpreted as adjectives and common nouns considered as proper names, most notably "wetlands" and "wastewater" occurring 3 and 5 times respectively. If the wrong POS is assigned, the words cannot be found in WordNet or the wrong synsets are assigned. In that case, wrong or no ontological statements are inserted for a word.

To analyze the recall in more detail, we looked at the most-frequent missed relations: patient (48) and done-by (30) (see Table 2).

Relation	Gold	%	System	Correct	R.	P.	Missed
destination-of	27	7.65%	17	6	22	35	21
use-of	4	1.13%	1	1	25	100	3
generic-location	11	3.12%	22	8	72	36	3
source-of	4	1.13%	10	1	25	10	3
instrument	2	0.57%	0	0	0	0	2
product-of	2	0.57%	0	0	0	0	2
part-of	1	0.28%	3	0	0	0	1
purpose-of	7	1.98%	9	3	42	33	4
patient	133	37.68%	195	85	63	43	48
path-of	1	0.28%	0	0	0	0	1
result-of	4	1.13%	7	0	0	0	4
participant	0	0.0%	3	0	0	0	0
has-state	32	9.07%	42	11	34	26	21
state-of	22	6.23%	25	11	50	44	11
done-by	52	14.73%	89	22	42	24	30
simple-cause-of	51	14.45%	125	26	50	20	25
Total	353	100%	548	174	49	31	179

Table 2: Generic processing with 261 profiles differentiated per relation

From the patient triplets, we missed 25% due to parser errors, among which wrong-POS, missed verb-particle combinations and multiwords. Another 15% of the patient triplets was not found because the parser does not provide detailed and reliable dependency information to distinguish between subjects and objects and the ontology does not distinguish sufficiently between events with participants that control the process (e.g. "to swim") and participants that do not (e.g. "to flow"). Remarkably, only 4% of the errors are due to a missing concept in WordNet or a wrong mapping of WordNet to the ontology. Another 4% could have been found by making more profiles.

In the case of the done-by relation, 30% of the missed relations are the result of parser errors (mainly coordination of NPs and VPs in which only one is marked as the head) and another 30% because the structures of simple-cause-of and done-by are the same and the ontology does not provide sufficient information on the events to distinguish.

Precision errors are mostly caused by the fact that patient, done-by and simple-cause-of are easily confused not only by the Kybots but also by humans. The patient relation performs slightly above average precision: 43% but done-by (24%), simple-cause-of (20%) and has-state (26%) are performing below average. Especially, the simple-cause-of relation is decreasing the overall precision since it represents 125 triplets (15%). The simple-cause-of relation applies to perdurants related to other perdurants. Due to the ambiguity in English of nouns to denote either an endurant or a perdurant, the system is likely to over-generate this relation. The reverse holds for the done-by relation. Both relations typically hold for the same structures such as nouns in subject position of a verb. Another common error that is related are cases such as *forest destroyed* and *houses built*. Since the parser does not provide information on the inflection of the verbs nor on passive/active form, the profile can only detect a noun+verb pattern and assigns a done-by relation where a patient relation should be assigned. Again, more information from the parser in the KAF representation can help here.

The main conclusion is that major improvement both in recall and precision can be achieved by better and more input from the linguistic preprocessing, by richer ontological information e.g. control of events, and by extending the number of profiles. Furthermore, precision could also be improved if we can resolve ambiguity between endurants and perdurants of nouns to distinguish for example done-by from simple-cause-of. It thus makes sense to consider the effect of WSD on the precision of the mining. This is discussed in the next section.

4.2.2 Effects of Word-Sense-Disambiguation

The generic processing considers all the possible meanings of the words and does not take the WSD into account.

To see the effect of the WSD, we implemented a filter on the Kybot output that selects interpretations with the highest WSD score for each word in the output that has multiple interpretations. By interpretation we mean: being either an event or a role or having different relations assigned. By excluding low scoring concepts only when there is a choice to be made, we hope to capture as much recall as possible and to gain precision. Note that the WSD scored a precision of 48% in the SemEval2010 task on domain specific WSD, which used documents from the same domain as KYOTO ((Agirre et al., 2010)). We set a threshold for eliminating relations in proportion to the maximum WSD scores of each word. The results are shown in Table 3. A threshold of 0 means that all interpretations are considered, a threshold of 100 means only the highest scoring interpretations.

We can see that there is a positive correlation between WSD threshold and precision, where precision increases from 32% to 39% using the highest WSD scores only. Recall drops from 49% to 42%. We get the optimal settings using a threshold for WSD of 60%. This gives an F-measure of 42%, for precision 38% and recall 46%.

WSD threshold	#triplets	#correct	P.	R.	F1
0	548	174	0.32	0.49	0.39
10	500	169	0.34	0.48	0.40
20	479	167	0.35	0.47	0.40
30	470	167	0.36	0.47	0.41
40	461	166	0.36	0.47	0.41
50	446	164	0.37	0.46	0.41
60	434	164	0.38	0.46	0.42
70	429	162	0.38	0.46	0.41
80	427	161	0.38	0.46	0.41
90	426	161	0.38	0.46	0.41
100	377	148	0.39	0.42	0.41
manual	364	141	0.39	0.40	0.39

Table 3: Generic processing with different WSD thresholds.

We also applied a manual disambiguation of the benchmark file. The results for the manually disambiguated file are shown in the last row. We can see that less triplets are generated (364) but close to the 100% WSD threshold (377). Remarkably, the precision is the same as for 100% WSD while recall is a bit less (40%). This shows that the errors of WSD apparently do not have a big impact on the extraction. For recall, it is thus better to use less perfect WSD. This is inline with the error analysis in the previous section, which showed that structural processing is more a problem than the mapping of the text to concepts.

We also checked the effect of WSD on the extraction of relevant events. Eliminating synsets through WSD did not show any effect. Precision remains the same (29%), and recall only drops slightly from 104% to 97% when we limit the events to 100% WSD threshold. In the case of manual WSD, we do get a much higher precision (49%) and a bit lower recall (83%). This clearly shows that the Kybots over-generate many events due to the event-object ambiguity of words in English. The fact that precision of the manually tagged file is much less than the recall, also suggest that relevance of extracted events is not considered by our system: even after manual (perfect) WSD, the system detects events that are not annotated. If we consider 49% of event detection as the upper limit here, which seems reasonable, we can say that the Kybots reach a precision of 60% of the upper limit in detecting events.

4.2.3 Effects of selecting best performing profiles

The profiles perform very differently in terms of recall and precision. We therefore derived the precision for each each profile, using the optimal WSD setting of 60% of the maximum score. We implemented a filter that checks every conflict across triplets. If two triplets involve the same events and roles but have a different relation, we choose the triplet generated by the higher scoring profile. The results are shown in Table 4. The first row of the table shows the results for a WSD threshold of 60% using 128 profiles. The remaining rows show the results when this 60% WSD output is post-filtered using profiles with precision scores 1, 5, 10, 25, 50 and 75.

	#profiles	#triplets	#correct	P.	R.	F1
All profiles	129	434	164	0.38	0.46	0.42
profiles 1%	104	332	147	0.44	0.42	0.43
profiles 5%	103	312	147	0.47	0.42	0.44
profiles 10%	103	312	147	0.47	0.42	0.44
profiles 25%	93	284	141	0.50	0.40	0.44
profiles 50%	76	219	115	0.53	0.33	0.40
profiles 75%	22	46	32	0.70	0.09	0.16

Table 4: Generic processing with WSD threshold of 60% and using best performing profiles.

We see a clear increase in precision and a drop in recall, as expected. However, we also see an increase in the F-measure from 41% to 44% using a subset of the profiles with higher precision. Using profiles with a precision score of at least 25%, we obtain a precision of 50% and a recall of 40%. With these settings, 90 profiles have been used compared to 129 profiles using just the WSD threshold of 60%. This shows that the set of profiles can be optimized for specific document collections by annotating a proportion of the collection that is representative and deriving a precision-score for the different profiles. Likewise, we can pair style of writing to the type of relation expressed.

If we compare these results with the manually annotated file in Table 3, we see that the best profiles have a much higher precision (50% against 39% manual) and the same recall. This again confirms that the challenge for getting more precision is in resolving the structural relations in the text rather than assigning better concepts through WSD.

5 Transferring Kybots to another language

An important aspect of the KYOTO system is the sharing of the central ontology and the possibility to extract semantic relations in different languages in a uniform way. To test the feasibility of sharing the same semantic backbone and transferring Kybot profiles, we carried out a transfer experiment from English to Dutch. We collected 93 Dutch documents on a Dutch estuary (the Westerschelde) and related topics. We created KAF files and applied WSD to these KAF file using the Dutch wordnet data.

To apply the profiles to the Dutch KAF documents, we need to apply the ontotagger program to the Dutch KAF. We created tables that match every Dutch synset to the English Base Concepts and to the ontology using the equivalence relations. We generated 145,189 Dutch synset to English Base Concept mappings (for comparison for English we have 114,477 mappings) and 326,667 Dutch synset to ontology mappings (186,383 for English). These ontotag tables were used to insert the ontological implications into the Dutch KAF files.

Next, we adapted the 261 English Kybot profiles to replace all English specific elements by Dutch. This mainly involved:

- replacing English prepositions and relative clause complementizers by Dutch equivalents;
- adapting the word order sequences for relative clauses in Dutch;
- adapting profiles that include adverbials, since they occur in different positions in Dutch;
- eliminating profiles for multiword compounds which mostly occur in Dutch as a one word compound;
- eliminating profiles for explicit English structures that express causal relations;

We kept all the ontological constraints exactly as they were for English. Only superficial syntactic properties were thus changed. It took us half-a-day to adapt the profiles for Dutch. From the original 261 English profiles, we obtained 134 Dutch profiles.

We ran the profiles on the 93 Dutch KAF files (42,697 word tokens) and 65 profiles generated output: 4,095 events and 6,862 roles. In terms of relations, we see a similar distribution as for English, as shown in Table 5. The patient relation is most frequent, followed by relations such as generic-location, has-state and done-by. We did a

Relation	#	%
destination-of	10	0,15%
patient	2067	30,12%
path-of	23	0,34%
has-state	1236	18,01%
generic-location	396	5,77%
state-of	748	10,90%
source-of	669	9,75%
done-by	792	11,54%
part-of	87	1,27%
simple-cause-of	573	8,35%
purpose-of	261	3,80%
Total	6,862	

Table 5: Relations extracted for Dutch documents.

preliminary inspection and the results look reasonable. For instance, two frequent words denoting events (the noun *toename* (increase) and the verb *stijgen* (increase)) appear to have sensible patients (*number*, *activity*, *consumption*, *pollution*, *trade*, *pressure*, *ground sea level*, *earth*).

6 Conclusions

We described an open platform for event-mining using wordnets and a central ontology that aims at maximizing the information extracted from text. The system uses a limited set of generic patterns with structural and ontological constraints on elements from the text. We have shown that wordnets can be used to map text to ontological classes and extract events and participants from text. Our error analysis showed that recall is mostly hampered by the structural complexity of the text and the incapability of the parser to handle this phenomenon. The knowledge resources, wordnet and the ontology, did not play a major role in recall. However, precision of the event relations is more affected by richness and quality of the semantics analysis. We have shown that WSD has a positive effect on the precision of the extracted relations and that precision can be further optimized by tuning the structural profiles to the genre of the target text. The system can be easily transferred to any language that has a wordnet connected to the English WordNet, as was shown for Dutch. In the future, we want to further improve recall and precision using richer event data and machine learning techniques and use the output for reconstruction of relations between events. We will also experiment with other parsers for English and Dutch to see the effect on the quality.

Acknowledgments

This work was been possible thanks to the support provided by KYOTO ICT-2007-211423, PATHS ICT-2010-270082 and KNOW2 TIN2009-14715-C04-04 projects.

References

E. Agirre and A. Soroa. 2009. Personalizing pagerank for word sense disambiguation. In *EACL*, pages 33–41.

Eneko Agirre, Montse Cuadros, German Rigau, and Aitor Soroa. 2010. Exploring knowledge bases for similarity. In Nicoletta Calzolari (Conference Chair), Khalid Choukri, Bente Maegaard, Joseph Mariani, Jan Odijk, Stelios Piperidis, Mike Rosner, and Daniel Tapias, editors, *Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10)*, Valletta, Malta, may. European Language Resources Association (ELRA).

W. E. Bosma, Piek Vossen, Aitor Soroa, German Rigau, Maurizio Tesconi, Andrea Marchetti, Monica Monachini, and Carlo Aliprandi. 2009. Kaf: a generic semantic annotation format. In *Proceedings of the GL2009 Workshop on Semantic Annotation*, Pisa, Italy.

C. Fellbaum. 1998. *WordNet: An Electronical Lexical Database*. The MIT Press, Cambridge, MA.

A. Gangemi, N. Guarino, C. Masolo, and A. Oltramari. 2003. Sweetening wordnet with dolce. *AI Mag.*, 24(3):13–24.

A. Hicks and A. Herold. 2009. Evaluating ontologies with rudify. In Jan L. G. Dietz, editor, *Proceedings of the 2nd International Conference on Knowledge Engineering and Ontology Development (KEOD'09)*, pages 5–12. INSTICC Press.

N. Ide and L.Romary. 2003. Outline of the international standard linguistic annotation framework. In *Proceedings of ACL'03 Workshop on Linguistic Annotation: Getting the Model*, pages 1–5.

R. Izquierdo, A. Suarez, and G. Rigau. 2007. Exploring the automatic selection of basic level concepts. In Galia Angelova et al., editor, *International Conference Recent Advances in Natural Language Processing*, pages 298–302, Borovets, Bulgaria.

K. Kaiser and S. Miksch. 2005. Information extraction. a survey. Technical report, Vienna University of Technology. Institute of Software Technology and Interactive Systems.

L. Peshkin and A. Pfeffer. 2003. Bayesian information extraction network. In *In Proc. of the 18th International Joint Conference on Artificial Intelligence*.

P. Vossen and G. Rigau. 2010. Division of semantic labor in the global wordnet grid. In *Proc. of Global WordNet Conference (GWC'2010). Mumbay, India*.