A Multilingual Predicate Matrix

Maddalen Lopez de Lacalle[◇]**, Egoitz Laparra**[⋆]**, Itziar Aldabe**[⋆]**, German Rigau**[⋆]
◇ Elhuyar Foundation
⋆ IXA group, University of the Basque Country, UPV/EHU
m.lopezdelacalle@elhuyar.com,{egoitz.laparra, itziar.aldabe, german.rigau}@ehu.eus

Abstract

This paper presents the Predicate Matrix 1.3, a lexical resource resulting from the integration of multiple sources of predicate information including FrameNet, VerbNet, PropBank and WordNet. This new version of the Predicate Matrix has been extended to cover nominal predicates by adding mappings to NomBank. Similarly, we have integrated resources in Spanish, Catalan and Basque. As a result, the Predicate Matrix 1.3 provides a multilingual lexicon to allow interoperable semantic analysis in multiple languages.

Keywords: Lexicon, Semantics, Multilinguality

1. Introduction

Predicate resources such as VerbNet (Kipper, 2005), FrameNet (Baker et al., 1997), PropBank (Palmer et al., 2005) and WordNet (Fellbaum, 1998) offer individually interesting characteristics not provided by their alternatives. Unfortunately, these semantic resources are developed independently and they are not fully integrated in a common framework. Thus, a common semantic infraestructure would allow the interoperability among all these lexicons. With this aim, we developed the Predicate Matrix (López de Lacalle et al., 2014b; López de Lacalle et al., 2014a; López de Lacalle et al., 2016 fc), a lexical resource resulting from the automatic integration of multiple sources of predicate information including FrameNet, VerbNet, PropBank and WordNet.

One of the few projects working on the integration of predicate information is SemLink (Palmer, 2009). The aim of this project is to link together different predicate resources via manual mappings. In contrast, the Predicate Matrix is built by automatic methods. Following different methods and techniques to automatically integrate different knowledge bases that contain predicate and role information, we have completed and extended the existing manual mappings provided by SemLink. In (López de Lacalle et al., 2016 fc), we demonstrate that these methods increase the number of mappings that are included in SemLink. Furthermore, our automatic methodology makes easier to maintain updated the set of mappings when new versions of the knowledge resources (each one developed independently) are released. As a result, in (López de Lacalle et al., 2016 fc), we present the Predicate Matrix 1.2 which integrates SemLink and a new set of mappings obtained by automatic methods. Section 3. presents the main steps to build the Predicate Matrix 1.2 as well as its main characteristics.

The semantic resources that are part of the Predicate Matrix 1.2 only contain an English verbal lexicon. But, if any other semantic resource is linked to any of the resources included in the Predicate Matrix the integration of the new resource can be done straightforwardly. In this paper we demonstrate this feature by extending the Predicate Matrix to English nominal predicates and to Spanish, Basque and Catalan languages (cf. Section 4.) and we also present the Predicate Matrix 1.3[1] that includes nominalizations and multilingual predicates.

In the case of the English nominal predicates, we use NomBank (Meyers et al., 2004). The projection to Spanish and Catalan is possible thanks to the AnCora (Taulé et al., 2008a) corpus and the AnCoraVerb (Aparicio et al., 2008) and AnCora-Nom semantic resources. Finally, the Basque Verb Index (BVI) (Estarrona et al., 2015) corpus-based lexicon is used in the case of Basque.

This paper is organized as follows. First, Section 2. includes a summary of the related work. Section 3. presents a brief description of the Predicate Matrix 1.2 with an overview of the methods developed to build the resource. In Section 4., we explain how we have extended the Predicate Matrix with nominal predicates and for multiple languages, and we present the resulting version 1.3. We finalize with some conclusions and future work in Section 5..

2. Related Work

Several previous works have been focused on the integration of resources targeted at knowledge about nouns and named entities. Well known examples are YAGO (Suchanek et al., 2007), Freebase (Bollacker et al., 2008), DBpedia (Bizer et al., 2009), BabelNet (Navigli and Ponzetto, 2010) or UBY (Gurevych et al., 2012). However, less attention has been paid to the integration of existing models for verbs and predicates (Burchardt et al., 2005; Fellbaum and Baker, 2013).

Regarding predicate information, SemLink (Palmer, 2009) has focused on mapping complementary lexical resources that associate semantic information to verbal predicates in a sentence. The resources integrated in SemLink vary in the detail and abstraction level of the encoded semantic information associated to each predicate. SemLink aims at unifying all these lexical resources. First by providing type-to-type mappings between the lexical units for each resource. Then, for each lexical unit, SemLink also supplies a mapping between the semantic roles of PropBank and VerbNet, as well as the roles of VerbNet and FrameNet. However, SemLink has some limitations. First, its coverage is still far from being complete (López de Lacalle et al., 2014b).

[1] http://adimen.si.ehu.es/web/PredicateMatrix

Second, the mappings between resources have been manually developed. A very costly process which is also not systematic. Our proposal is to define automatic methods for mapping different semantic resources containing predicate information in order to allow semantic interoperability between them.

3. Predicate Matrix 1.2

In (López de Lacalle et al., 2016 fc), we present a proposal which defines a set of automatic methods for mapping the semantic knowledge included in WordNet, VerbNet, PropBank and FrameNet. The integration of predicate information is performed first at a lexical level, and second at a role level. After studying different settings for each method using SemLink as a gold-standard for evaluation, we prioritize precision over recall so that we give preference to more reliable aligments. Then, we build the Predicate Matrix 1.2 integrating the mappings automatically obtained with those existing in SemLink.

	Precision	Recall	F-score
WN-VN	85.3	85.3	85.3
WN-FN	82.9	81.8	82.4
WN-PB	71.3	58.0	64.0
PB-FN	89.8	52.4	66.2

Table 1: Results of the methods used to obtain the automatic mapping between predicates of Predicate Matrix 1.2. WN: WordNet; FN: FrameNet; VN: VerbNet; PB: PropBank.

All the mappings obtained at the lexical level are based on graph-based Word Sense Disambiguation (WSD) algorithms. More specifically, the lexical mappings from WordNet to FrameNet and also to VerbNet are obtained by applying WSD algorithms to semantically coherent groupings of verbal entries whereas the lexical mappings from WordNet to PropBank are obtained by applying the WSD to a corpus annotated with PropBank predicates. A corpus-based approach crossing different annotations is used to create automatic predicate mappings between FrameNet and PropBank. Table 1 shows the results of the evaluation of these methods.[2] We have not created new mappings between PropBank and VerbNet because PropBank already offers this information and its coverage is quite complete.

As it happens with the lexical mappings, PropBank also offers quite complete role mappings between PropBank and VerbNet. Thus, we concentrate our efforts on finding new role mappings between FrameNet and VerbNet and between FrameNet and PropBank. The mappings between FrameNet frame-elements and VerbNet thematic roles are obtained following a three-step process whereas the same corpus-based approach used previously for predicates is applied to automatically create new role mappings between FrameNet and PropBank. The performances of these methods is showed in Table 2.[2]

	Precision	Recall	F-score
VN-FN	87.1	84.7	85.9
PB-FN	75.0	41.2	53.2

Table 2: Results of the methods used to obtain the automatic mapping between roles of Predicate Matrix 1.2. WN: WordNet; FN: FrameNet; VN: VerbNet; PB: PropBank.

Tables 3 and 4 show the differences between SemLink and the Predicate Matrix in terms of mappings between lexicons (Table 3) and roles (Table 4). Thanks to the methodology proposed for creating automatic mappings between lexical entries and roles, the resource obtained is much larger than SemLink. Note that the Predicate Matrix 1.2 arises from the union of SemLink and the set of mappings obtained by our automatic methods.

4. Including nominalizations and multilingual predicates

The semantic resources that are part of the Predicate Matrix 1.2 only contain English predicates in their verbal form. However, we can easily extend the Predicate Matrix to nominal predicates or to other languages if any other semantic resource in linked to some of the resources included in the Predicate Matrix. In this section, we describe the approach we have followed to obtain a multilingual Predicate Matrix that also includes nominal predicates. For this purpose, we have made use of the mappings existing between the following resources:

- English nominal predicates:
  PropBank(PB)-NomBank(NB)
- Spanish verbal predicates:
  PropBank(PB)-SpanishAnCoraVerb(SAV)
- Spanish nominal predicates:
  SpanishAnCoraVerb(SAV)-SpanishAnCoraNom(SAN)
- Catalan verbal predicates:
  PropBank(PB)-CatalanAnCoraVerb(CAV)
- Basque verbal predicates:
  PropBank(PB)/VerbNet(VN)-Basque Verb Index(BVI)

Table 5 contains the number of mappings listed above both for the lexical entries and roles. Note that the case of Basque is special. Unlike the rest of languages, where both predicates and roles are mapped between the same resources, for Basque, the predicates of the Basque Verb Index are mapped to PropBank and the roles are linked to VerbNet.

4.1. Multilingual Predicate Matrix

The strategy to project the Predicate Matrix to new languages is very simple, if there is a resource in that language linked to any of the resource included in the Predicate Matrix. This is the case of AnCora (Taulé et al., 2008b),[3] a multilevel corpus that includes both for Spanish

[2] (López de Lacalle et al., 2016 fc) provides complete details on the methodology and evaluation for building the Predicate Matrix version 1.2

[3] http://clic.ub.edu/corpus/ancora

	WN-VN	WN-PB	WN-FN	VN-PB	VN-FN	PB-FN
SemLink	7,665	5,489	4,851	4,503	3,709	2,562
Predicate Matrix	10,832	9,516	8,583	4,947	5,462	4,163

Table 3: Differences between SemLink and the Predicate Matrix 1.2: Mappings between lexicons. WN: WordNet; FN: FrameNet; VN: VerbNet; PB: PropBank.

	PB-VN	FN-VN	FN-PB
SemLink	9,950	6,934	4,384
Predicate Matrix	11,749	14,258	14,195

Table 4: Differences between SemLink and the Predicate Matrix: Mappings between roles. WN: WordNet; FN: FrameNet; VN: VerbNet; PB: PropBank.

and Catalan, annotations of lemmatization, syntactic constituents, WordNet senses, coreference, named entities and also semantic roles. AnCora also develops a semantic resource called AnCoraVerb (Aparicio et al., 2008) that contains Spanish and Catalan verbal predicates and their corresponding arguments structures (see Table 6).

AnCoraVerb is based on PropBank and both resources are linked by a wide set of mappings called AncoraNet. The Spanish predicate *"verb.vender.1.default"* shown in Table 7 is for instance mapped to the English predicate *"sell.01"* and the Catalan predicate *"verb.dialogar.1.default"* is mapped to the English *"speak.01"*.

Unless AncoraNet states otherwise, the correspondence between the arguments in AncoraVerb and PropBank is direct. For the Spanish predicate *"verb.vender.1.default"*, the arguments *"arg0"*, *"arg1"* and *"arg2"* correspond respectively to the arguments *"0"*, *"1"* and *"2"* of the English predicate *"sell.01"*. The Catalan predicate *"verb.dialogar.1.default"* has its *"arg0"* linked directly to PropBank argument *"0"*, but AncoraNet also explicitly establishes (cf. Table 7) that the *"arg1"* corresponds to the argument *"2"* of PropBank. We use these alignments to duplicate the entries in the Predicate Matrix. For example, all the mappings involving the argument *"0"* of the predicate *"sell.01"* are projected to argument *"arg0"* of the Spanish predicate *"verb.vender.1.default"*.

In the case of Basque, the mappings are defined in a slightly different way. The predicates of the Basque Verb Index are mapped to PropBank and the roles are linked to VerbNet. For example, as shown in Table 8, the Basque predicate *"ordain_1"* is linked to the English predicate *"sell.01"* of PropBank and the arguments *"0"*, *"1"* and *"2"* of *"ordain_1"* are mapped to their corresponding VerbNet roles, *"agent"*, *"theme"* and *"recipient"* respectively.

Nevertheless, the projection to Basque can be performed because both PropBank and VerbNet are part of the Predicate Matrix.

4.2. Nominal Predicate Matrix

Extending the Predicate Matrix to nominal predicates follows the same strategy as the one previously explained for the projection to a new language, provided the existence of semantic resources containing the argument structures for the nominalizations of the verbal predicates. In the case of English predicates, this knowledge can be obtained from NomBank (Meyers et al., 2004)[4] that includes nominalizations of the PropBank predicates, like *"sale.01"* aligned to the source verbal predicate *"sell.01"* (see Table 9).

For Spanish, AnCora also includes the nominalizations of its verbal predicates in a resource called AnCora-Nom. For example, *"venta.1.default"*, the nominalization of the predicate *"vender.1.default"*, is described in AnCora-Nom as shown in Table 10.

Once again, unless these resources state otherwise, the correspondence between the arguments of the verbal and nominal predicates is direct. Hence, the lines we showed previously for *"sell.01"* can be replicated for its nominalization.

4.3. Predicate Matrix 1.3

Following the strategy presented in the previous section, we have built the Predicate Matrix 1.3, that includes nominalizations and multilingual predicates. As a result of including NomBank, AnCoraVerb, AnCoraNom and the Basque Verb Index into the Predicate Matrix, we have obtained new mappings between these resources and VerbNet, FrameNet and WordNet. In Tables 11 and 12, we show the number of resulting mappings we obtain for Predicate Matrix v1.3.

Table 13 shows some examples of the resulting records in the new version of the Predicate Matrix. Note that we have defined an identifier to distinguish between lines for English, Basque, Spanish and Catalan predicates, and between their verbal and nominal forms. This identifier is based on PropBank, AnCora and the Basque Verb Index predicates and arguments, and is composed of 4 fields: language, form, predicate and argument. For example, according to Table 13, the line that correspond to the argument *"1"* of the English nominal predicate *"sale.01"* is identified by *"id:eng id:n id:sale.01 id:1"*. Similarly, the corresponding line for argument *"arg0"* of the Spanish verbal predicate *"vender.1.default"* is identified by *"id:spa id:v id:vender.1.default id:arg0"*, as shown in Table 13. The Catalan and Basque lines are indexed by *"id:cat"* and *"id:eus"* respectively. Establishing such identifiers allows us to maintain the whole Predicate Matrix for all the languages in the same file.

4.3.1. Additional results

Extending the Predicate Matrix to languages other than English produces some additional outcomes. First, as shown in previous sections, we have linked WordNet to several resources by mapping predicates (verbal and nominal) to the interlingual index (ILI) of the Multilingual Central Repository (MCR). The ILI is a unique identifier that

[4] http://nlp.cs.nyu.edu/meyers/NomBank.html

	PB-NB	PB-SAV	SAV-SAN	PB-CAV	PB-BVI	VN-BVI
lexicon	3,494	7,966	2,966	6,493	506	-
roles	10,307	22,544	9,132	18,660	-	1,408

Table 5: Number of mappings between different resources. PB: PropBank; VN: VerbNet; NB:NomBank; SAV: Spanish AnCora-Verbs; SAN: Spanish AnCora-Nouns; CAV: Catalan AnCora-Verbs; BVI: Basque Verb Index.

```
<lexentry lemma="vender" lng="es" type="verb">
  <sense id="1">
    <frame lss="A32.ditransitive-patient-benefactive"
    default="yes" type="default">
      <argument argument="arg0" function="suj"
      thematicrole="agt"/>
      <argument argument="arg1" function="cd"
      thematicrole="pat"/>
      <argument argument="arg2" function="ci"
      thematicrole="ben">
      ...
<lexentry lemma="dialogar" lng="ca" type="verb">
  <sense id="1">
    <frame lss="A22.transitive-agentive-theme"
    default="yes" type="default">
      <argument argument="arg0" function="suj"
      thematicrole="agt"/>
      <argument argument="arg1" function="creg"
      thematicrole="tem">
      ...
```

Table 6: Examples of argument structures defined in AnCoraVerb for the predicates *"vender.1.default"* and *"dialogar.1.default"*.

```
<link ancoralexid="verb.vender.1.default"
propbankid="sell.01">
</link>

<link ancoralexid="verb.dialogar.1.default"
propbankid="speak.01">
  <arglink ancoralexarg="arg1" propbankarg="2"/>
</link>
```

Table 7: Examples of mappings between AnCoraVerb and PropBank predicates.

```
<aditz aditza="saldu" zenb="222" >
  <adiera zenb="1" >
    <ordain zenb="1" adiera="sell_01" >
      <arg zenb="0" rol="agent"
      eadbrol="abiapuntua" >
        <case grammcase="erg"/>
      </arg>
      <arg zenb="1" rol="theme"
      eadbrol="gaia">
        <case grammcase="abs"/>
      </arg>
      <arg zenb="2" rol="recipient"
      eadbrol="helburua">
        <case grammcase="dat"/>
      </arg>
    </ordain>
  </adiera>
</aditzh>
```

Table 8: Example of the argument structure defined in Basque Verb Index for the predicate *"ordain 1"* and its corresponding mappings to PropBank and VerbNet.

```
<roleset id="sale.01" name="commerce: seller"
source="verb-sell.01" vncls="13.1-1">
  <roles>
    <role descr="seller" n="0">
      <vnrole vncls="13.1-1" vntheta="Agent"/>
    </role>
    <role descr="thing sold" n="1">
      <vnrole vncls="13.1-1" vntheta="Theme"/>
    </role>
    <role descr="buyer" n="2">
      <vnrole vncls="13.1-1" vntheta="Recipient"/>
    </role>
```

Table 9: Examples of argument structures defined in NomBank for the predicate *"sale.01"*.

joints synsets of different languages with the same meaning. For example, the ili *ili-30-00007739-v* represents the English synset *eng-30-00007739-v blink_1 wink_3 nictitate_1 nictate_1* and the Spanish synset *spa-00007739-v pestañear_1*. However, in the new multilingual Predicate Matrix this Spanish synset also have associated the Spanish word senses *parpadear* and *guiñar*. Thus, the integration of additional senses for Spanish, Catalan and Basque also produces as a side effect the enrishment of the Spanish, Catalan and Basque wordnets integrated into the MCR. Table 14 presents the total number of new senses aligned to the different wordnets. Interestingly, some additional word senses are also created for the English WordNet. We plan to include this new word sense aligments in future releases of the MCR.

4.3.2. The Predicate Matrix in NewsReader

The Predicate Matrix 1.3 is part of a multilingual event detection system implemented within the NewsReader project[5] (Vossen et al., 2014). The NewsReader project develops advanced technology to process daily news streams in 4 languages (Agerri et al., 2016), extracting what happened, when and where it happened and who was involved. With this purpose, the event detection system is a pipeline which contains a set of modules to perform various NLP tasks. Among others, the system has a semantic role labeling module that automatically annotates semantic information based on PropBank. Thanks to the Predicate Matrix, our pipelines also obtain the equivalent annotations in

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

```
<lexentry lemma=”venta” lng=”es” origin=”deverbal”
type=”noun”>
    <sense originlemma=”vender” id=”1” cousin=”no”
    originlink=”verb.vender.1” denotation=”result”
    lexicalized=”no” wordnetsynset=”16:00721968”>
        <frame appearsinplural=”yes” type=”default”>
            <argument argument=”arg0”
            thematicrole=”agt”>
            </argument>
            <argument argument=”arg1”
            foundincorporated=”yes” thematicrole=”pat”>
            </argument>
            <argument argument=”arg2”
            thematicrole=”ben”>
            </argument>
```

Table 10: Description of the nominal predicate *“venta.1.default”* in AnCora-Nom.

	PB	VN	FN	WN
NB	2,963	3,923	3,911	7,430
SAV	6,745	9,092	8,777	15,310
SAN	4,469	6,190	6,157	10,747
CAV	5,529	7,567	7,347	13,109
BVI	415	652	745	1,330

Table 11: Number of lexicon Mappings in PMv1.3. NB: NomBank; SAV: SpanishAnCoraVerb; SAN: SpanishAnCoraNom; CAV: CatalanAnCoraVerb; BVI: Basque Verb Index; WN: WordNet; FN: FrameNet; VN: VerbNet; PB: PropBank.

	PB	VN	FN
NB	7,699	9,699	10,351
SAV	17,152	19,173	20,296
SAN	11,752	13,177	14,439
CAV	14,307	16,174	17,204
BVI	1,048	1,275	1,629

Table 12: Number of role Mappings in PMv1.3. NB: NomBank; SAV: SpanishAnCoraVerb; SAN: SpanishAnCoraNom; CAV: CatalanAnCoraVerb; BVI: Basque Verb Index; WN: WordNet; FN: FrameNet; VN: VerbNet; PB: PropBank.

id:eng id:n id:sale.01 id:1 vn:give-13.1 vn:Theme wn:ili-30-02244956-v fn:Commerce_sell fn:Goods pb:sell.01 pb:1
id:spa id:v id:vender.1.default id:arg0 vn:give-13.1 vn:Agent wn:ili-30-02244956-v fn:Commerce_sell fn:Seller pb:sell.01 pb:0
id:spa id:n id:venta.1.default id:arg2 vn:give-13.1 vn:Recipient wn:ili-30-02244956-v fn:Commerce_sell fn:Buyer pb:sell.01 pb:2
id:cat id:v id:dialogar.1.default id:arg0 vn:talk-37.11 vn:Agent wn:ili-30-00941990-v fn:Chatting fn:Interlocutor_1 pb:speak.01 pb:0
id:eus id:v id:saldu.1 id:1 vn:give-13.1 vn:Theme wn:ili-30-02242464-v fn:Commerce_sell fn:Goods pb:sell.01 pb:1

Table 13: Some examples of mappings in the Predicate Matrix 1.3

FrameNet and VerbNet. Our pipelines guarantee interoperability across language and predicate resources by integrating the Predicate Matrix within the SRL modules. As the Predicate Matrix gathers multilingual knowledge bases that contain predicate and semantic role information, it is possible to know which lexical-semantic units refer to the same events or roles.

Figure 1 provides the output of our SRL module for the English sentence *Steve Jobs gave his annual opening speech to the WWDC at Moscone Center, on Monday*. Our SRL module first processes the sentence providing predicates and role annotations from PropBank. Now, as PropBank is integrated into the Predicate Matrix, our SRL module can also obtain the corresponding predicate classes and roles for the rest of the predicate resources. Thus, *Steve Jobs* identified as *A0* role of the nominal predicate *speach.01* corresponds to the *Communicator* role of a *Communication* frame according to FrameNet. Thus, thanks to the Predicate Matrix, predicates and roles from PropBank appear also aligned to the rest of resources.

This also applies across languages thanks to the multilingual Predicate Matrix. Thus, when processing the Spanish sentence by our SRL module, we obtain the same language-independent semantic representation as the one obtained from the English sentence. Now, using this new multilingual version of the Predicate Matrix we can also obtain the same event representation for a Spanish translation of the English sentence *Steve Jobs ofreció el lunes su conferencia inaugural de la WWDC en el Moscone Center*. That is, that *Steve Jobs* corresponds to the *Communicator* role of a *Communication* frame according to FrameNet. This same process was performed with Dutch and Italian role resources.

Furthermore, in NewsReader the events and their participants are also aligned to the Event and Situation Ontology (ESO) (Segers et al., 2016). ESO formalizes with preconditions and post-conditions events and roles and reuses existing resources such as WordNet, SUMO (Niles and Pease, 2001) and FrameNet.

5. Conclusion

Building large and rich predicate models takes a great deal of expensive manual effort. Furthermore, the same effort should be invested for each different language. Predicate resources such as VerbNet, FrameNet, PropBank and WordNet offer individually interesting characteristics not provided by their alternatives. Unfortunately, these semantic resources are developed independently and they are not fully integrated in a common framework. Thus, a common semantic infraestructure would allow the interoperability among all these resources.

One of the few projects working on the integration of the predicate information is SemLink (Palmer, 2009). However, the mappings of this resource has been developed by manual means and only cover verbal predicates. Moreover, SemLink only takes into account English lexicon.

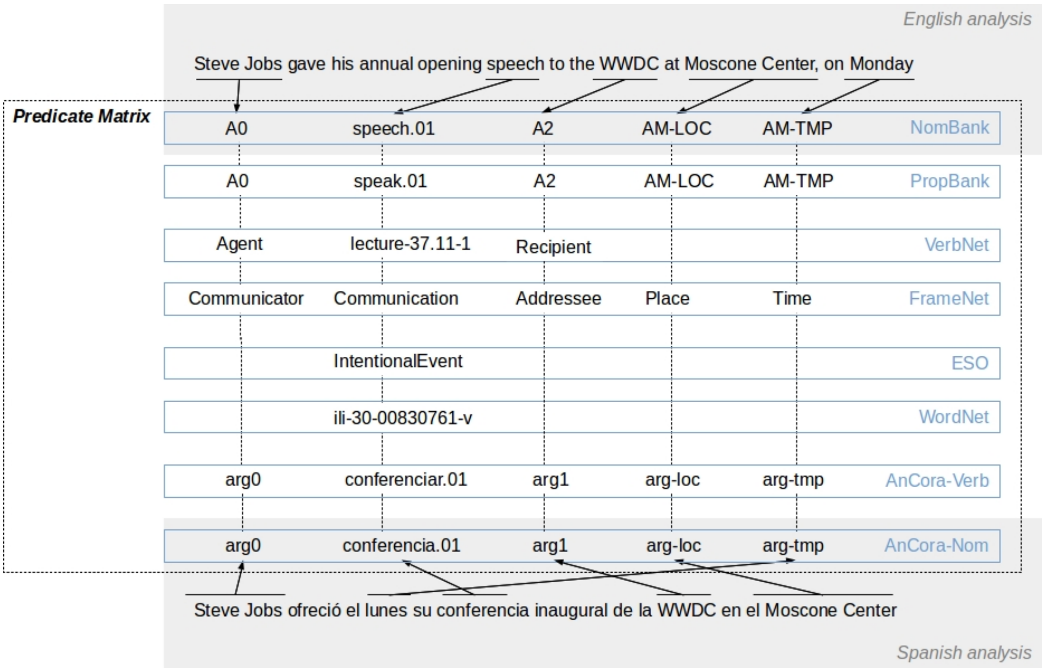

Figure 1: Cross-lingual and semantic interoperability provided by the Predicate Matrix

	new word senses
English WN	53
Spanish WN	6,092
Catalan WN	5,182
Basque WN	855

Table 14: Number of new word senses created for the different wordnets.

In this work, we have developed the Predicate Matrix 1.3, a new lexical-semantic resource resulting from the union of mappings obtained by automatic methods and SemLink that has been extended to Spanish, Catalan and Basque. In this new version, we have included NomBank (Meyers et al., 2004) and we have integrated Ancora (Taulé et al., 2008a) and the Basque Verb Index (BVI) (Estarrona et al., 2015). This process allows to project the predicate information of the Predicate Matrix to Spanish, Catalan and Basque.

With the new Predicate Matrix, we expect to provide a multilingual interoperable predicate lexicon. As future work, we want to perform some experiments to evaluate the Predicate Matrix indirectly. Our idea is to apply the new mappings across languages in order to obtain improvements in some NLP tasks such as semantic role labeling for those languages with low resources, as is the case of the Catalan or Basque.

6. Acknowledgments

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

7. Bibliographical References

Agerri, R., Aldabe, I., Laparra, E., Rigau, G., Fokkens, A., Huijgen, P., van Erp, M., Bevia, R. I., Vossen, P., Minard, A.-L., and Magnini, B. (2016). Multilingual event detection using the newsreader pipelines. In *Procedings of the 1st Workshop on Cross-Platform Text Mining and Natural Language Processing Interoperability (INTEROP2016) at the 10th Language Resources and Evaluation Conference (LREC2016)*, Portorož, Slovenia.

Aparicio, J., Taulé, M., and Martí, M. A. (2008). Ancora-verb: A lexical resource for the semantic annotation of corpora. In *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*, Marrakech, Morocco.

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

Bizer, C., Lehmann, J., Kobilarov, G., Auer, S., Becker, C., Cyganiak, R., and Hellmann, S. (2009). Dbpedia-a crystallization point for the web of data. *Web Semantics: Science, Services and Agents on the World Wide Web*, 7(3):154–165.

Bollacker, K., Evans, C., Paritosh, P., Sturge, T., and Taylor, J. (2008). Freebase: a collaboratively created graph database for structuring human knowledge. In *Proceedings of the 2008 ACM SIGMOD international conference on Management of data*, pages 1247–1250. ACM.

Burchardt, A., Erk, K., and Frank, A. (2005). A WordNet Detour to FrameNet. In *Proceedings of the GLDV 2005 GermaNet II Workshop*, pages 408–421, Bonn, Germany.

Estarrona, A., Aldezabal, I., and Díaz de Ilarraza, A. (2015). A methodology for the semiautomatic annotation of epec-rolsem, a basque corpus labeled at predicate level following the propbank-verbnet model. *Digital Scholarship in the Humanities*, pages 1–23.

Fellbaum, C. and Baker, C. F. (2013). Comparing and harmonizing different verb classifications in light of a semantic annotation task. *Linguistics*, 51(4):707–728.

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

Gurevych, I., Eckle-Kohler, J., Hartmann, S., Matuschek, M., Meyer, C. M., and Wirth, C. (2012). Uby: A large-scale unified lexical-semantic resource based on lmf. In *Proceedings of EACL*, pages 580–590.

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

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

López de Lacalle, M., Laparra, E., and Rigau, G. (2014b). First steps towards a predicate matrix. In *Proceedings of the 7th Global WordNet Conference (GWC2014)*, Tartu, Estonia, January 25-29.

López de Lacalle, M., Laparra, E., Aldabe, I., and Rigau, G. (2016 (f.c.)). Predicate matrix: Automatically extending the interoperability between predicate resources. *Language Resources and Evaluation*.

Meyers, A., Reeves, R., Macleod, C., Szekely, R., Zielinska, V., Young, B., and Grishman, R. (2004). The nombank project: An interim report. In *In Proceedings of the NAACL/HLT Workshop on Frontiers in Corpus Annotation*.

Navigli, R. and Ponzetto, S. P. (2010). Babelnet: Building a very large multilingual semantic network. In *Proceedings of the 48th annual meeting of ACL*, pages 216–225.

Niles, I. and Pease, A. (2001). Towards a standard upper ontology. In *Proceedings of the 2nd International Conference on Formal Ontology in Information Systems (FOIS-2001)*, pages 17–19. Chris Welty and Barry Smith, eds.

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

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

Segers, R., Rospocher, M., Vossen, P., Laparra, E., Rigau, G., and Minard, A.-L. (2016). The event and implied situation ontology (eso): Application and evaluation. In *Proceedings of the 10th Language Resources and Evaluation Conference (LREC2016)*, Portorož, Slovenia, May 23-28.

Suchanek, F. M., Kasneci, G., and Weikum, G. (2007). Yago: A Core of Semantic Knowledge. In *WWW conference*, New York, NY, USA. ACM Press.

Taulé, M., Martí, M. A., and Recasens, M. (2008a). Ancora: Multilevel annotated corpora for catalan and spanish. In *LREC*.

Taulé, M., Martí, M. A., and Recasens, M. (2008b). Ancora: Multilevel annotated corpora for catalan and spanish. In *LREC*.

Vossen, P., Rigau, G., Serafini, L., Stouten, P., Irving, F., and Hage, W. R. V. (2014). Newsreader: recording history from daily news streams. In *Proceedings of the 9th Language Resources and Evaluation Conference (LREC2014)*, Reykjavik, Iceland, May 26-31.