

First steps towards a Predicate Matrix

Maddalen López de Lacalle

IXA group, UPV/EHU
Donostia, Spain

maddalen.lopezdelacalle@ehu.es

Egoitz Laparra

IXA group, UPV/EHU
Donostia, Spain

egoitz.laparra@ehu.es

German Rigau

IXA group, UPV/EHU
Donostia, Spain

german.rigau@ehu.es

Abstract

This paper presents the first steps towards building the Predicate Matrix, a new lexical resource resulting from the integration of multiple sources of predicate information including FrameNet (Baker et al., 1997), VerbNet (Kipper, 2005), PropBank (Palmer et al., 2005) and WordNet (Fellbaum, 1998). By using the Predicate Matrix, we expect to provide a more robust interoperable lexicon by discovering and solving inherent inconsistencies among the resources. Moreover, we plan to extend the coverage of current predicate resources (by including from WordNet morphologically related nominal and verbal concepts), to enrich WordNet with predicate information, and possibly to extend predicate information to languages other than English (by exploiting the local wordnets aligned to the English WordNet).

1 Introduction

Predicate models such as FrameNet (Baker et al., 1997), VerbNet (Kipper, 2005) or PropBank (Palmer et al., 2005) are core resources in most advanced NLP tasks, such as Question Answering, Textual Entailment or Information Extraction. Most of the systems with Natural Language Understanding capabilities require a large and precise amount of semantic knowledge at the predicate-argument level. This type of knowledge allows to identify the underlying typical participants of a particular event independently of its realization in the text. Thus, using these models, different linguistic phenomena expressing the same event, such as active/passive transformations, verb alternations, nominalizations, implicit realizations can be harmonized into a common semantic representation. In fact, lately, several systems have been developed for shallow semantic parsing an explicit and implicit semantic role labeling using these resources (Erk and Pado, 2004), (Shi and Mihalcea,

2005), (Giuglea and Moschitti, 2006), (Laparra and Rigau, 2013).

However, building large and rich enough predicate models for broad-coverage semantic processing takes a great deal of expensive manual effort involving large research groups during long periods of development. In fact, the coverage of currently available predicate-argument resources is still far from complete. For example, (Burchardt et al., 2005) or (Shen and Lapata, 2007) indicate the limited coverage of FrameNet as one of the main problems of this resource. In fact, FrameNet1.5 covers around 10,000 lexical-units while for instance, WordNet3.0 contains more than 150,000 words. Furthermore, the same effort should be invested for each different language (Subirats and Petrucci, 2003). Moreover, most previous research efforts on the integration of resources targeted at knowledge about nouns and named entities rather than predicate knowledge. 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).

Following the line of previous works (Shi and Mihalcea, 2005), (Burchardt et al., 2005), (Johansson and Nugues, 2007), (Pennacchiotti et al., 2008), (Cao et al., 2008), (Tonelli and Pianta, 2009), (Laparra et al., 2010), we will also focus on the integration of predicate information. We start from the basis of SemLink (Palmer, 2009). SemLink aimed to connect together different predicate resources such as FrameNet (Baker et al., 1997), VerbNet (Kipper, 2005), PropBank (Palmer et al., 2005) and WordNet (Fellbaum, 1998). However, its coverage is still far from complete.

The Predicate Matrix, the resource resulting from the work presented in this paper, will allow to extend the coverage of current predicate resources (by including from WordNet closely related nominal and verbal concepts), to discover in-

herent inconsistencies among the resources, to enrich WordNet with predicate information, and possibly to extend predicate information to languages other than English (by exploiting the local wordnets aligned to the English WordNet). Moreover, the Predicate Matrix uses WordNet as a central resource. In that way, each row (or line) in the matrix presents partial predicate information related to a particular WordNet word sense.

First, as SemLink takes VerbNet as the central resource, we present a complete study of the coverage of the mappings between each resource included in SemLink to VerbNet. We describe the coverage and gaps of these mappings with respect to the lexical entries and the role structures of each resource. Second, we exploit WordNet to propose straightforward methods to discover inconsistencies among the resources as well as to extend their coverage towards a more complete and robust predicate lexicon.

2 Sources of Predicate information

As a starting point for building the Predicate Matrix, we consider the sources of predicate knowledge connected to SemLink.

SemLink¹ (Palmer, 2009) is a project whose aim is to link together different predicate resources establishing a set of mappings. These mappings make it possible to combine the different information provided by the different lexical resources for tasks such as inferencing, consistency checking, interoperable semantic role labelling, etc. Currently, SemLink provides partial mappings to VerbNet (Kipper, 2005), PropBank (Palmer et al., 2005), FrameNet (Baker et al., 1997) and WordNet (Fellbaum, 1998).

FrameNet² (Baker et al., 1997) is a very rich semantic resource that contains descriptions and corpus annotations of English words following the paradigm of Frame Semantics (Fillmore, 1976). In frame semantics, a Frame corresponds to a scenario that involves the interaction of a set of typical participants, playing a particular role in the scenario. FrameNet groups words or lexical-units (LUs hereinafter) into coherent semantic classes or frames, and each frame is further characterized by a list of participants or frame-elements (FEs hereinafter). Different senses for a word are represented in FrameNet by assigning different frames.

PropBank³ (Palmer et al., 2005) aims to provide a wide corpus annotated with information about semantic propositions, including relations between the predicates and their arguments. PropBank also contains a description of the frame structures, called framesets, of each sense of every verb that belong to its lexicon. Unlike other similar resources, as FrameNet, PropBank defines the arguments, or roles, of each verb individually. In consequence, it becomes a hard task obtaining a generalization of the frame structures over the verbs.

VerbNet⁴ (Kipper, 2005) is a hierarchical domain-independent broad-coverage verb lexicon for English. VerbNet is organized into verb classes extending (Levin, 1993) classes through refinement and addition of subclasses to achieve syntactic and semantic coherence among members of a class. Each verb class in VerbNet is completely described by thematic-roles, selectional restrictions on the arguments, and frames consisting of a syntactic description and semantic predicates.

WordNet⁵ (Fellbaum, 1998) is by far the most widely-used knowledge base. In fact, WordNet is being used world-wide for anchoring different types of semantic knowledge including wordnets for languages other than English (Gonzalez-Agirre et al., 2012a). It contains manually coded information about English nouns, verbs, adjectives and adverbs and is organized around the notion of a *synset*. A synset is a set of words with the same part-of-speech that can be interchanged in a certain context. For example, *<learn, study, read, take>* form a synset because they can be used to refer to the same concept. A synset is often further described by a gloss, in this case: *"be a student of a certain subject"* and by explicit semantic relations to other synsets. Each synset represents a concept that are related with an large number of semantic relations, including hypernymy/hyponymy, meronymy/holonymy, antonymy, entailment, etc.

Obviously, we can also exploit the existing SemLink mappings to aid semi-automatic or fully automatic extensions of the current mapping coverage, in order to increase the overall overlapping.

¹<http://verbs.colorado.edu/semlink/>

²<http://framenet.icsi.berkeley.edu/>

³<http://verbs.colorado.edu/~mpalmer/projects/ace.html>

⁴<http://verbs.colorado.edu/~mpalmer/projects/verbnet.html>

⁵<http://wordnet.princeton.edu/>

3 SemLink coverage

As SemLink uses VerbNet as the central resource, we present a complete study of the coverage of the mappings between each resource included in SemLink to VerbNet.

3.1 WordNet and VerbNet alignment

Although VerbNet is one of largest verb lexicons available it does not reach the coverage of the verbal part of WordNet. While WordNet contains **25,047** different verb senses there are just **6,293** predicates in VerbNet classes. This means that the mapping between both resources is, obviously, incomplete. Specifically there are **18,559** senses of WordNet, corresponding to **9,995** different lemmas, that have not been assigned to any VerbNet predicate. Many of these cases appear because of the distinct granularities of both resources. In fact **6,120** WordNet senses (corresponding to **2,099** lemmas) that are not mapped to VerbNet belong to lemmas that have at least another WordNet sense properly mapped to VerbNet. For instance, Table 1 shows the mapping between the verb *drown* in WordNet and VerbNet. Note that only two of the five WordNet senses are assigned to VerbNet.

WordNet sense	VerbNet	
	member	class
drown%2:30:00::	drown	40.7
drown%2:31:00::	drown	42.2
	drown	40.7
drown%2:30:02::	-	-
drown%2:35:00::	-	-
drown%2:42:00::	-	-

Table 1: WordNet to VerbNet alignment for *drown_v*

The rest of missing senses correspond to those cases where the lemma does not exist in the VerbNet lexicon (**7,320** lemmas and **11,201** senses). For example the verb *abort* does not appear in VerbNet since its three WordNet senses are not part of SemLink. The remaining cases (**1,443** WordNet senses and **576** lemmas) correspond to lemmas that exist in both resources but there is no sense mapping between them. For instance, there is no mapping between the WordNet sense *harm%2:29:00::* and the VerbNet verb that belongs to the class 31-1.

Moreover, SemLink does not provide mappings to WordNet senses for **1,077** VerbNet predicates. **304** of these VerbNet predicates share the same lemma with some other VerbNet sense that is al-

ready mapped to a WordNet sense. This is the case of *reveal* as shown in Table 2.

VerbNet		WordNet
member	class	sense
reveal	29.2	reveal%2:32:00
reveal	37.7	reveal%2:32:00
reveal	37.10	reveal%2:32:00
reveal	48.1.2	-
reveal	78	-

Table 2: VerbNet to WordNet alignment for *reveal_v*

From the rest of missing members, **574** correspond to those cases the lemma of the predicate also exists in WordNet (like the example of *harm* explained previously). Finally, there are only **199** verb senses in VerbNet whose lemmas do not exist in WordNet. For example: *africanize_v*, *backfill_v* or *carbonify_v*.

3.2 PropBank and VerbNet alignment

The mapping between PropBank and VerbNet introduces additional complexity to the comparison of both resources. In this case, aligning the lexicon means that the arguments of the PropBank predicates must be aligned to the VerbNet thematic-roles.

First, regarding the lexicon mapping, once again, the differences in the coverages of the resources impede to obtain a complete alignment. From the **6,181** different PropBank predicates (comprising **4,552** lemmas), just **3,558** have their corresponding VerbNet predicate in SemLink. That is, **2,623** PropBank predicates have no correspondances to VerbNet. However, all the lemmas of PropBank are contained within the VerbNet lexicon. This means that for each one of the **2,623** missing predicates from PropBank there exists at least another predicate with the same lemma that is mapped to VerbNet. That is the case of the PropBank predicate *abandon.02*, shown in Table 3.

PropBank predicate	VerbNet	
	member	class
accept.01	accept	13.5.2
	accept	29.2-1-1
	accept	77
abandon.01	abandon	51.2
abandon.02	-	-

Table 3: PropBank to VerbNet alignments for *accept_v* and *abandon_v*

On the other hand, we found that the num-

ber of VerbNet predicates that are not aligned to PropBank is smaller than the number of PropBank predicates not aligned to VerbNet. That is, up to **4,736** of the **6,293** VerbNet predicates are aligned to PropBank while only **1,557** VerbNet predicates are not aligned to PropBank. Moreover, **298** of these VerbNet predicates do not exist in the PropBank lexicon, for instance *arrogate_v*, *deconstruct_v*, *mewl_v* or *sprint_v*. Finally, there are **312** VerbNet predicates whose lemmas (**265** in total) are actually part of the PropBank lexicon but there is no alignment for them. For example, the predicate *offload_v* of the VerbNet class *wipe_manner-10.4.1* is not connected to the PropBank predicate *offload.01*. Table 4 shows some alignments from VerbNet to PropBank.

VerbNet		PropBank
member	class	predicate
laugh	40.2	laugh.01
flow	47.2	flow.01
flow	48.1.1	-

Table 4: VerbNet to PropBank alignments for *laugh_v* and *flow_v*

Regarding the PropBank arguments and the VerbNet thematic-roles, **7,915** out of **15,871** arguments from PropBank⁶ are mapped to a thematic-role from VerbNet⁷. That is, around a half of the total PropBank arguments, leaving out the remaining **7,956** arguments. From the opposite point of view, **9,682** out of **17,382** thematic-roles from VerbNet are included in the SemLink mapping. This means that **7,700** thematic-roles are not aligned to any PropBank argument. Table 5 contains some examples of existing and also missing mappings between PropBank arguments and VerbNet thematic-roles.

3.3 FrameNet and VerbNet alignment

The alignment between FrameNet and VerbNet proves to be very incomplete. For example, only **1,730** lexical-units from FrameNet⁸ are aligned to, at least, one VerbNet predicate⁹. This number represents only 16% out of the total **10,195** lexical-units of FrameNet. Table 6 presents some align-

⁶Arguments of particular PropBank predicates. For instance, Arg0 of *paint.01*.

⁷Thematic-roles of particular VerbNet predicates. For instance, Agent of *paint_v*.

⁸Lexical-units of particular FrameNet frames. For instance, *sell_v* from the frame *Commerce_sell*.

⁹Predicate of a particular VerbNet class. For instance, *sell_v* from 13.1-1 VerbNet class.

VerbNet		PropBank	
predicate	them-role	predicate	argument
paint 9.9	Agent	paint.01	A0
paint 9.9	Destination	paint.01	A1
paint 9.9	Theme	paint.01	A2
plant 9.7	Agent	-	-
plant 9.7	Destination	-	-
plant 9.7	Theme	-	-
abandon 51.2	Theme	abandon.01	A0
-	-	abandon.01	A1
-	-	abandon.01	A2

Table 5: Some alignments between VerbNet thematic-roles and PropBank arguments

ments between VerbNet predicates and FrameNet lexical-units.

VerbNet		FrameNet	
class	member	frame	lexical-unit
13.1-1	sell	Commerce_sell	sell.v
13.5.1	buy	Commerce_buy	buy.v
53.1-1	delay	Hindering	delay.v
53.1-1	delay	Change_event_time	-
13.5.3	employ	Employing	-
105	employ	Using	-

Table 6: Some alignments between VerbNet predicates and FrameNet lexical-units (LUs)

SemLink also includes the alignment between the roles of both resources. However, unlike PropBank the roles of FrameNet, that are called frame-elements, are defined at frame-level and not at predicate level. Therefore, the mapping of the VerbNet thematic-roles and the frame-elements of FrameNet is defined between VerbNet classes and FrameNet frames. Table 7 presents an example of the alignment of some roles from both resources for the VerbNet class 54.1. A VerbNet predicate member of this class has been aligned to the WordNet sense *total%2:42:00*.

VerbNet		FrameNet	
class	thematic-role	frame	frame-element
54.1	Agent	Adding_up	Cognizer
54.1	Theme	Adding_up	Numbers
54.1	Value	Adding_up	Result

Table 7: Some alignments between VerbNet thematic-roles and FrameNet frame-elements (FEs)

Once again, the mapping between VerbNet and FrameNet presents significant gaps and miss-

matches. For instance, just **825** of the **7,124** frame-elements of FrameNet¹⁰ are linked to a VerbNet thematic-role. That is, **88%** of the frame-elements from FrameNet are not aligned to any VerbNet thematic-role. Moreover, only **262** frames out of **795** have at least one frame-element aligned to a VerbNet thematic-role. That is, just a few frames are used in the mapping. However, it also seems that, at a class level, most of the VerbNet thematic-roles appear to be aligned to at least one frame-element. VerbNet covers **787** different thematic-roles¹¹. From these, **541** appear to be aligned to a FrameNet frame-element. In other words, it seems that just **246** thematic-roles are missing from the mapping provided by SemLink. Table 8 presents some class level alignments between VerbNet thematic-roles and FrameNet frame-elements (FEs).

VerbNet		FrameNet	
class	thematic-role	frame	frame-element
10.10	Agent	-	-
10.10	Attribute	-	-
10.10	Source	-	-
10.10	Theme	-	-
48.3	Theme	Catastrophe	Undesirable_Event
48.3	Location	Catastrophe	Place
48.3	Location	Catastrophe	Time
-	-	Catastrophe	Cause
-	-	Catastrophe	Circumstances
-	-	Catastrophe	Degree
-	-	Catastrophe	Manner
-	-	Catastrophe	Undergoer
-	-	Addiction	Addict
-	-	Addiction	Addictant
-	-	Addiction	Compeller
-	-	Addiction	Degree
-	-	Addiction	State

Table 8: Some alignments between VerbNet thematic-roles and FrameNet frame-elements (FEs)

4 Using WordNet to cross-check predicate information

In this section, as a proof-of-concept, we will show a simple way to exploit WordNet for validating the predicate information appearing in SemLink. We apply a very simple method to check the consistency of VerbNet. Consider the following WordNet synset <*understand*, *read*, *interpret*,

¹⁰Frame-elements of a particular FrameNet frame. For instance, the frame-element Cognizer for the *Adding up* frame

¹¹Role of a particular VerbNet class. For instance, Agent of VerbNet class 10.10

translate> with the gloss “make sense of a language” and the example sentences “*She understands French; Can you read Greek?*”. As synonyms, these verbs denote the same concept and are interchangeable in many contexts. However, in SemLink *read*%2:31:04 appears aligned with the VerbNet class *learn-14-1*¹² while one of its synonyms *understand*%2:31:03 appears aligned to the VerbNet class *comprehend-87.2*¹³. Moreover, the thematic-roles of both classes are different. *Learn-14-1* has the following thematic-roles Agent (with semantic type [+animate]), Topic and Source while *comprehend-87.2* has Experiencer (with semantic type [+animate or +organization]), Attribute and Stimulus. Are both sets of thematic-roles compatible? Complementary? Is one of them incorrect? Should we joint them? Maybe is the alignment incorrect? Is perhaps the synset definition?

Following with this example, the VerbNet predicate *understand*_v has no connection to FrameNet, but its VerbNet class *comprehend-87.2-1* has some other verbal predicates aligned to FrameNet. For instance, *apprehend*_v, *comprehend*_v and *grasp*_v are linked to the *Grasp*¹⁴ FrameNet frame. Among the lexical-units corresponding to the *Grasp* frame it appears also the verbal predicate *understand*_v. This means that possibly, this verbal predicate should also be aligned to the FrameNet frame *Grasp*. The core frame-elements (roles) of this frame are Cognizer (with semantic type Sentient), Faculty and Phenomenon. Is this set of roles compatible with the previous ones?

5 Using WordNet to extend SemLink

As we have seen in Section 3, the mapping between the different sources of predicate information is far from being complete. However, the existing alignments also offer a very interesting source of information to be systematically exploited. In fact, we are devising a number of simple automatic methods to extend SemLink by exploiting simple properties from WordNet. As a proof-of-concept, we present in this section two very simple approaches to extend the coverage of the mapping between VerbNet predicates and

¹²<http://verbs.colorado.edu/verb-index/vn/learn-14.php#learn-14-1>

¹³<http://verbs.colorado.edu/verb-index/vn/comprehend-87.2.php#comprehend-87.2-1>

¹⁴<https://framenet2.icsi.berkeley.edu/fnReports/data/frame/Grasp.xml>

WordNet senses. Moreover, we also plan to use additional semantic resources that use WordNet as a backbone. For instance, exploiting those knowledge resources integrated into the Multilingual Central Repository¹⁵ (MCR) (Atserias et al., 2004; Gonzalez-Agirre et al., 2012a) to extend automatically the alignment of the different sources of predicate information (VerbNet, PropBank, FrameNet and WordNet). Following the line of previous works, in order to assign more WordNet verb senses to VerbNet predicates, we also plan to apply more sophisticated word-sense disambiguation algorithms to semantically coherent groups of predicates (Laparra et al., 2010).

5.1 VerbNet Monosemous predicates

Monosemous verbs from WordNet can be directly assigned to VerbNet predicates still without a WordNet alignment. This very simple strategy solves **240** alignments. In this way, VerbNet predicates such as *divulge_v*, *exhume_v*, *mutate_v* or *upload_v* obtain a corresponding WordNet word sense¹⁶. Remember that only **576** lemmas from VerbNet were not aligned to WordNet.

5.2 WordNet synonyms

A very straightforward method to extend the mapping between WordNet and VerbNet consists on including synonyms of already aligned WordNet senses as new members of the corresponding VerbNet class. Obviously, this method expects that WordNet synonyms share the same predicate information. For instance, the predicate *desert_v* member of the VerbNet class *leave-51.2-1* appears to be assigned to *desert%2:31:00* WordNet verbal sense. In WordNet, this word sense also has three synonyms, *abandon_v*, *forsake_v* and *desolate_v*. Obviously, these three verbal senses can also be assigned to the same VerbNet class. This simple approach can create up to **5,075** new members of VerbNet classes (corresponding to **4,616** different WordNet word senses). For instance, Table 9 presents two productive examples. Moreover, applying this method **103** VerbNet predicates without mapping to WordNet in SemLink are aligned to a WordNet word sense.

¹⁵<http://adimen.si.ehu.es/MCR>

¹⁶Obviously, these alignments can be considered just as suggestions to be revised later on manually.

VerbNet	WordNet	New
leave-51.2.1	desert%2:31:00	abandon%2:31:00:: forsake%2:31:00:: desolate%2:31:00::
remove-10.1	retract%2:32:00	abjure%2:32:00:: recant%2:32:00:: forswear%2:32:00:: resile%2:32:00::

Table 9: New WordNet senses aligned to VerbNet

6 A first version of the Predicate Matrix

We already produced a preliminary version of the Predicate Matrix¹⁷. The original SemLink in a Predicate Matrix form resulted in 36,174 rows (corresponding to 6,556 WordNet word senses). By applying the synonyms method described in the previous section the Predicate Matrix extended to 69,508 rows (10,984 WordNet word senses). Finally, by applying the monosemous method, the Predicate Matrix further extended to 70,391 rows (11,146 WordNet word senses).

Table 10 presents a full example of the information that is currently available in the Predicate Matrix including the new mappings obtained by the methods described in the previous section. Each row of this Table represents the mapping of a role over the different resources and includes all the aligned knowledge about its corresponding verb. The Table presents the cases obtained originally from SemLink, denoted as *SEMLINK*, and the cases inferred following the methods explained previously, identified as *SYNONYMS* or *MONOSEMIC* depending on the case. The Table also includes the following fields: the lemma and the class in VerbNet, the sense of the verb in WordNet, the thematic-role in VerbNet, the Frame of FrameNet, the corresponding lexical-entry and frame-element of FrameNet, the predicate in PropBank and its argument, the offset of the sense in WordNet and the knowledge associated with that sense in the MCR, such as the Adimen-SUMO (Álviz et al., 2012) and the new WordNet domain aligned to WordNet 3.0 (González-Agirre et al., 2012b) features as well as the Base Level Concept (Izquierdo et al., 2007) of the sense. Finally, each line also includes the frequency and the number of relations of the WordNet word sense.

¹⁷<http://adimen.si.ehu.es/web/PredicateMatrix>

7 Conclusions and future work

We are now producing and studying initial versions of the Predicate Matrix by exploiting Sem-Link and applying very simple methods to extend and validate its content. By developing more advanced versions of the Predicate Matrix, we expect to provide a more robust and interoperable predicate lexicon. We plan to discover and solve inherent inconsistencies among the integrated resources. Moreover, we plan to extend the coverage of current predicate resources (by including from WordNet morphologically related nominal and verbal concepts, by exploiting also FrameNet information, etc.), to enrich WordNet with predicate information, and possibly to extend predicate information to languages other than English (by exploiting the local wordnets aligned to the English WordNet) and predicate information from other languages. For instance, the Ancora Spanish corpus and lexicon (Taulé et al., 2008).

Acknowledgment

We are grateful to the anonymous reviewers for their insightful comments. This work has been partially funded by SKaTer (TIN2012-38584-C06-02), OpeNER (FP7-ICT-2011-SME-DCL-296451) and NewsReader (FP7-ICT-2011-8-316404), as well as the READERS project with the financial support of MINECO, ANR (convention ANR-12-CHRI-0004-03) and EPSRC (EP/K017845/1) in the framework of ERA-NET CHIST-ERA (UE FP7/2007-2013).

References

- Javier Álvez, Paqui Lucio, and German Rigau. 2012. Adimen-sumo: Reengineering an ontology for first-order reasoning. *International Journal on Semantic Web and Information Systems (IJSWIS)*, 8(4):80–116.
- Jordi Atserias, Luís Villarejo, German Rigau, Eneko Agirre, John Carroll, Bernardo Magnini, and Piek Vossen. 2004. The meaning multilingual central repository. In *Proceedings of GWC*, Brno, Czech Republic.
- Collin F. Baker, Charles J. Fillmore, and John B. Lowe. 1997. The berkeley framenet project. In *COLING/ACL'98*, Montreal, Canada.
- Christian Bizer, Jens Lehmann, Georgi Kobilarov, Sören Auer, Christian Becker, Richard Cyganiak, and Sebastian Hellmann. 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.
- Kurt Bollacker, Colin Evans, Praveen Paritosh, Tim Sturge, and Jamie Taylor. 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.
- Aljoscha Burchardt, Katrin Erk, and Anette Frank. 2005. A WordNet Detour to FrameNet. In *Proceedings of the GLDV 2005 GermaNet II Workshop*, pages 408–421, Bonn, Germany.
- Diego De Cao, Danilo Croce, Marco Pennacchiotti, and Roberto Basili. 2008. Combining word sense and usage for modeling frame semantics. In *Proceedings of The Symposium on Semantics in Systems for Text Processing (STEP 2008)*, Venice, Italy.
- Katrin Erk and Sebastian Pado. 2004. A powerful and versatile xml format for representing role-semantic annotation. In *Proceedings of LREC-2004*, Lisbon.
- Christiane Fellbaum, editor. 1998. *WordNet. An Electronic Lexical Database*. The MIT Press.
- Charles J. Fillmore. 1976. Frame semantics and the nature of language. In *Annals of the New York Academy of Sciences: Conference on the Origin and Development of Language and Speech*, volume 280, pages 20–32, New York.
- Ana-Maria Giuglea and Alessandro Moschitti. 2006. Semantic role labeling via framenet, verbnet and propbank. In *Proceedings of COLING-ACL 2006*, pages 929–936, Morristown, NJ, USA. ACL.
- Aitor Gonzalez-Agirre, Egoitz Laparra, and German Rigau. 2012a. Multilingual central repository version 3.0. In *LREC*, pages 2525–2529.
- Aitor González-Agirre, German Rigau, and Mauro Castillo. 2012b. A graph-based method to improve wordnet domains. In *CICLING*, pages 17–28. Springer.
- Iryna Gurevych, Judith Eckle-Kohler, Silvana Hartmann, Michael Matuschek, Christian M Meyer, and Christian Wirth. 2012. Uby: A large-scale unified lexical-semantic resource based on lmf. In *Proceedings of EACL*, pages 580–590.
- Rubén Izquierdo, Armando Suárez, and German Rigau. 2007. Exploring the automatic selection of basic level concepts. In *Proceedings of RANLP*, volume 7. Citeseer.
- Richard Johansson and Pierre Nugues. 2007. Using WordNet to extend FrameNet coverage. In *Proceedings of the Workshop on Building Frame-semantic Resources for Scandinavian and Baltic Languages, at NODALIDA*, Tartu, Estonia, May 24.

VERSION	VN.LEMMA	VN.CLASS	WN.SENSE	VN.THEME/ROLE	FN.FRAME	FN.LEXENT	FN.SPEAKER	FN.ROLESET	PB.ARG	MCR.HOFFICE	MCR.SUMO	MCR.Domain	WN.BLC	WN.SENSE/FREC	WN.SYNSE/REL_NUM
SEMLINK	stutter	37.3	stutter%2:32:00	Agent	Communication_manner	565	Addresser	stutter.01	0	ih-30-00981544-v	Communication	factotum	stutter%2:32:00	2	006
SEMLINK	stutter	37.3	stutter%2:32:00	Recipient	Communication_manner	565	Message	stutter.01	2	ih-30-00981544-v	Communication	factotum	stutter%2:32:00	2	006
SEMLINK	stutter	37.3	stutter%2:32:00	Topic	Communication_manner	565	Topic	stutter.01	1	ih-30-00981544-v	Communication	factotum	stutter%2:32:00	2	006
SEMLINK	stammer	37.3	stammer%2:32:00	Agent	Communication_manner	564	Addresser	stutter.01	0	ih-30-00981544-v	Communication	factotum	stutter%2:32:00	6	006
SEMLINK	stammer	37.3	stammer%2:32:00	Recipient	Communication_manner	564	Message	stutter.01	2	ih-30-00981544-v	Communication	factotum	stutter%2:32:00	6	006
SEMLINK	stammer	37.3	stammer%2:32:00	Topic	Communication_manner	564	Topic	stutter.01	1	ih-30-00981544-v	Communication	factotum	stutter%2:32:00	6	006
SYNONYMS	bumble	37.3	bumble%2:32:00	Agent	Communication_manner	565	Addresser	stutter.01	0	ih-30-00981544-v	Communication	factotum	stutter%2:32:00	2	006
SYNONYMS	bumble	37.3	bumble%2:32:00	Recipient	Communication_manner	565	Message	stutter.01	2	ih-30-00981544-v	Communication	factotum	stutter%2:32:00	2	006
SYNONYMS	bumble	37.3	bumble%2:32:00	Topic	Communication_manner	565	Topic	stutter.01	1	ih-30-00981544-v	Communication	factotum	stutter%2:32:00	2	006
SYNONYMS	bumble	37.3	bumble%2:32:00	Agent	Communication_manner	564	Addresser	stutter.01	0	ih-30-00981544-v	Communication	factotum	stutter%2:32:00	6	006
SYNONYMS	bumble	37.3	bumble%2:32:00	Recipient	Communication_manner	564	Message	stutter.01	2	ih-30-00981544-v	Communication	factotum	stutter%2:32:00	6	006
SYNONYMS	bumble	37.3	bumble%2:32:00	Topic	Communication_manner	564	Topic	stutter.01	1	ih-30-00981544-v	Communication	factotum	stutter%2:32:00	6	006
SYNONYMS	falter	37.3	falter%2:32:00	Agent	Communication_manner	565	Addresser	stutter.01	0	ih-30-00981544-v	Communication	factotum	stutter%2:32:00	2	006
SYNONYMS	falter	37.3	falter%2:32:00	Recipient	Communication_manner	565	Message	stutter.01	2	ih-30-00981544-v	Communication	factotum	stutter%2:32:00	2	006
SYNONYMS	falter	37.3	falter%2:32:00	Topic	Communication_manner	565	Topic	stutter.01	1	ih-30-00981544-v	Communication	factotum	stutter%2:32:00	2	006
SYNONYMS	falter	37.3	falter%2:32:00	Agent	Communication_manner	564	Addresser	stutter.01	0	ih-30-00981544-v	Communication	factotum	stutter%2:32:00	6	006
SYNONYMS	falter	37.3	falter%2:32:00	Recipient	Communication_manner	564	Message	stutter.01	2	ih-30-00981544-v	Communication	factotum	stutter%2:32:00	6	006
SYNONYMS	falter	37.3	falter%2:32:00	Topic	Communication_manner	564	Topic	stutter.01	1	ih-30-00981544-v	Communication	factotum	stutter%2:32:00	6	006
SEMLINK	squeak	37.3	squeak%2:39:00	Agent	Communication_noise	612	Addresser	squeak.01	0	ih-30-02171664-v	SoundAttribute	factotum	squeak%2:39:00	3	007
SEMLINK	squeak	37.3	squeak%2:39:00	Recipient	Communication_noise	612	Message	squeak.01	1	ih-30-02171664-v	SoundAttribute	factotum	squeak%2:39:00	3	007
SEMLINK	squeak	37.3	squeak%2:39:00	Topic	Communication_noise	612	Topic	squeak.01	1	ih-30-02171664-v	SoundAttribute	factotum	squeak%2:39:00	3	007
SEMLINK	screech	37.3	screech%2:39:00	Agent	Communication_noise	604	Addresser	screech.03	0	ih-30-02171664-v	SoundAttribute	factotum	screech%2:39:00	5	007
SEMLINK	screech	37.3	screech%2:39:00	Recipient	Communication_noise	604	Message	screech.03	2	ih-30-02171664-v	SoundAttribute	factotum	screech%2:39:00	5	007
SEMLINK	screech	37.3	screech%2:39:00	Topic	Communication_noise	604	Topic	screech.03	1	ih-30-02171664-v	SoundAttribute	factotum	screech%2:39:00	5	007
SYNONYMS	scream	37.3	scream%2:39:00	Agent	Communication_noise	612	Addresser	squeak.01	0	ih-30-02171664-v	SoundAttribute	factotum	scream%2:39:00	5	007
SYNONYMS	scream	37.3	scream%2:39:00	Recipient	Communication_noise	612	Message	squeak.01	1	ih-30-02171664-v	SoundAttribute	factotum	scream%2:39:00	5	007
SYNONYMS	scream	37.3	scream%2:39:00	Topic	Communication_noise	612	Topic	squeak.01	1	ih-30-02171664-v	SoundAttribute	factotum	scream%2:39:00	5	007
SYNONYMS	scream	37.3	scream%2:39:00	Agent	Communication_noise	604	Addresser	screech.03	2	ih-30-02171664-v	SoundAttribute	factotum	scream%2:39:00	3	007
SYNONYMS	scream	37.3	scream%2:39:00	Recipient	Communication_noise	604	Message	screech.03	1	ih-30-02171664-v	SoundAttribute	factotum	scream%2:39:00	3	007
SYNONYMS	scream	37.3	scream%2:39:00	Topic	Communication_noise	604	Topic	screech.03	1	ih-30-02171664-v	SoundAttribute	factotum	scream%2:39:00	3	007
SYNONYMS	scream	37.3	scream%2:39:00	Agent	Communication_noise	604	Addresser	screech.03	0	ih-30-02171664-v	SoundAttribute	factotum	scream%2:39:00	5	007
SYNONYMS	scream	37.3	scream%2:39:00	Recipient	Communication_noise	604	Message	screech.03	2	ih-30-02171664-v	SoundAttribute	factotum	scream%2:39:00	5	007
SYNONYMS	scream	37.3	scream%2:39:00	Topic	Communication_noise	604	Topic	screech.03	1	ih-30-02171664-v	SoundAttribute	factotum	scream%2:39:00	5	007
SYNONYMS	white	37.3	white%2:39:00	Agent	Communication_noise	612	Addresser	squeak.01	0	ih-30-02171664-v	SoundAttribute	factotum	white%2:39:00	3	007
SYNONYMS	white	37.3	white%2:39:00	Recipient	Communication_noise	612	Message	squeak.01	1	ih-30-02171664-v	SoundAttribute	factotum	white%2:39:00	3	007
SYNONYMS	white	37.3	white%2:39:00	Topic	Communication_noise	612	Topic	squeak.01	1	ih-30-02171664-v	SoundAttribute	factotum	white%2:39:00	3	007
SYNONYMS	white	37.3	white%2:39:00	Agent	Communication_noise	604	Addresser	screech.05	0	ih-30-02171664-v	SoundAttribute	factotum	white%2:39:00	5	007
SYNONYMS	white	37.3	white%2:39:00	Recipient	Communication_noise	604	Message	screech.05	2	ih-30-02171664-v	SoundAttribute	factotum	white%2:39:00	5	007
SYNONYMS	white	37.3	white%2:39:00	Topic	Communication_noise	604	Topic	screech.05	1	ih-30-02171664-v	SoundAttribute	factotum	white%2:39:00	5	007
MONOSEMIC	hatter	37.3	hatter%2:52:00	Agent	Communication_manner	547	Addresser	hatter.01	0	ih-30-01038666-v	Communication	factotum	hatter%2:52:00	5	007
MONOSEMIC	hatter	37.3	hatter%2:52:00	Recipient	Communication_manner	547	Message	hatter.01	1	ih-30-01038666-v	Communication	factotum	hatter%2:52:00	5	007
MONOSEMIC	hatter	37.3	hatter%2:52:00	Topic	Communication_manner	547	Topic	hatter.01	1	ih-30-01038666-v	Communication	factotum	hatter%2:52:00	5	007
MONOSEMIC	hatter	37.3	hatter%2:52:00	Agent	Communication_manner	547	Addresser	hatter.01	0	ih-30-01038666-v	Communication	factotum	hatter%2:52:00	1	011
MONOSEMIC	hatter	37.3	hatter%2:52:00	Recipient	Communication_manner	547	Message	hatter.01	1	ih-30-01038666-v	Communication	factotum	hatter%2:52:00	1	011
MONOSEMIC	hatter	37.3	hatter%2:52:00	Topic	Communication_manner	547	Topic	hatter.01	1	ih-30-01038666-v	Communication	factotum	hatter%2:52:00	1	011
MONOSEMIC	prattle	37.3	prattle%2:32:00	Agent	Communication_manner	547	Addresser	hatter.01	0	ih-30-01038666-v	Communication	factotum	prattle%2:32:00	1	015
MONOSEMIC	prattle	37.3	prattle%2:32:00	Recipient	Communication_manner	547	Message	hatter.01	1	ih-30-01038666-v	Communication	factotum	prattle%2:32:00	1	015
MONOSEMIC	prattle	37.3	prattle%2:32:00	Topic	Communication_manner	547	Topic	hatter.01	1	ih-30-01038666-v	Communication	factotum	prattle%2:32:00	1	015
MONOSEMIC	prattle	37.3	prattle%2:32:00	Agent	Communication_manner	551	Addresser	jabber.01	0	ih-30-01051956-v	Communication	factotum	prattle%2:32:00	2	011
MONOSEMIC	prattle	37.3	prattle%2:32:00	Recipient	Communication_manner	551	Message	jabber.01	1	ih-30-01051956-v	Communication	factotum	prattle%2:32:00	2	011
MONOSEMIC	prattle	37.3	prattle%2:32:00	Topic	Communication_manner	551	Topic	jabber.01	1	ih-30-01051956-v	Communication	factotum	prattle%2:32:00	2	011
MONOSEMIC	rant	37.3	rant%2:32:00	Agent	Communication_manner	551	Addresser	jabber.01	0	ih-30-01051956-v	Communication	factotum	rant%2:32:00	2	011
MONOSEMIC	rant	37.3	rant%2:32:00	Recipient	Communication_manner	551	Message	jabber.01	2	ih-30-01051956-v	Communication	factotum	rant%2:32:00	2	011

Table 10: Partial view of the current content of the Predicate Matrix

- Karen Kipper. 2005. *VerbNet: A broad-coverage, comprehensive verb lexicon*. Ph.D. thesis, University of Pennsylvania.
- Egoitz Laparra and German Rigau. 2013. Impar: A deterministic algorithm for implicit semantic role labelling. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (ACL 2013)*, pages 33–41.
- Egoitz Laparra, German Rigau, and Montse Cuadros. 2010. Exploring the integration of wordnet and framenet. In *Proceedings of the 5th Global WordNet Conference (GWC 2010), Mumbai, India*.
- Beth Levin. 1993. *English verb classes and alternations: A preliminary investigation*, volume 348. University of Chicago press Chicago.
- Roberto Navigli and Simone Paolo Ponzetto. 2010. Babelnet: Building a very large multilingual semantic network. In *Proceedings of the 48th annual meeting of ACL*, pages 216–225.
- Martha Palmer, Daniel Gildea, and Paul Kingsbury. 2005. The proposition bank: An annotated corpus of semantic roles. *Computational Linguistics*, 31(1):71–106, March.
- Martha Palmer. 2009. Semlink: Linking propbank, verbnet and framenet. In *Proceedings of the Generative Lexicon Conference*, pages 9–15.
- Marco Pennacchiotti, Diego De Cao, Roberto Basili, Danilo Croce, and Michael Roth. 2008. Automatic induction of FrameNet lexical units. In *Proceedings of EMNLP*.
- Dan Shen and Mirella Lapata. 2007. Using semantic roles to improve question answering. In *Proceedings of the Joint Conference on (EMNLP-CoNLL)*, pages 12–21.
- Lei Shi and Rada Mihalcea. 2005. Putting pieces together: Combining framenet, verbnet and wordnet for robust semantic parsing. In *Proceedings of CILing, Mexico*.
- Carlos Subirats and Miriam R.L. Petruck. 2003. Surprise: Spanish framenet! In *Proceedings of the International Congress of Linguists, Praga*.
- Fabian M. Suchanek, Gjergji Kasneci, and Gerhard Weikum. 2007. Yago: A Core of Semantic Knowledge. In *WWW conference*, New York, NY, USA. ACM Press.
- Mariona Taulé, Maria Antònia Martí, and Marta Recasens. 2008. Ancora: Multilevel annotated corpora for catalan and spanish. In *LREC*.
- Sara Tonelli and Emanuele Pianta. 2009. A novel approach to mapping framenet lexical units to wordnet synsets. In *Proceedings of IWCS-8, Tilburg, The Netherlands*.