Exploring the Integration of WordNet and FrameNet

Egoitz Laparra
IXA group, UPV/EHU
Donostia, Spain
egoitz.laparra@ehu.com

German Rigau
IXA group, UPV/EHU
Donostia, Spain
german.rigau@ehu.com

Montse Cuadros
TALP Research Center, UPC
Barcelona, Spain
cuadros@lsi.upc.edu

Abstract

This paper presents a novel automatic approach to partially integrate FrameNet and WordNet. In that way we expect to extend FrameNet coverage, to enrich WordNet with frame semantic information and possibly to extend FrameNet to languages other than English. The method uses a knowledge-based Word Sense Disambiguation algorithm for matching the FrameNet lexical units to WordNet synsets. Specifically, we exploit a graph-based Word Sense Disambiguation algorithm that uses a large-scale knowledge-base derived from existing resources. We have developed and tested additional versions of this algorithm showing substantial improvements over state-of-the-art results. Finally, we show some examples and figures of the resulting semantic resource.

1 Introduction

Predicate models such as FrameNet (Baker et al., 1998), 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 and nominalizations can be harmonized into a common semantic representation. In fact, lately, several systems have been developed for shallow semantic parsing and semantic role labeling using these resources (Erk and Pado, 2004), (Shi and Mihalcea, 2005), (Giuglea and Moschitti, 2006).

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. Thus, the coverage of currently available predicate-argument resources is still unsatisfactory. 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, FrameNet1.3 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 Petruck, 2003). 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), we empirically study a novel approach to partially integrate FrameNet (Baker et al., 1998) and WordNet (Fellbaum, 1998). The method relies on the use of a knowledge-based Word Sense Disambiguation (WSD) algorithm that uses a large-scale graph of concepts derived from WordNet (Fellbaum, 1998) and eXtented WordNet (Mihalcea and Moldovan, 2001). The WSD algorithm is applied to coherent groupings of words belonging to the same frame. In that way we expect to extend the coverage of FrameNet (by including from WordNet closely related concepts), to enrich WordNet with frame semantic information (by porting frame information to Wordnet) and possibly to extend FrameNet to languages other than English (by exploiting local wordnets aligned to the English WordNet).

WordNet [1] (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 word-

[1] http://wordnet.princeton.edu/

nets for languages other than English (Atserias et al., 2004), domain knowledge (Magnini and Cavaglià, 2000) or ontologies like SUMO (Niles and Pease, 2001) or the EuroWordNet Top Concept Ontology (Álvez et al., 2008). 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, <*student*, *pupil*, *educatee*> 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: "a learner who is enrolled in an educational institution" 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.

FrameNet [2] (Baker et al., 1998) 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 lexical elements (LEs hereinafter). Different senses for a word are represented in FrameNet by assigning different frames.

Currently, FrameNet represents more than 10,000 LUs and 825 frames. More than 6,100 of these LUs also provide linguistically annotated corpus examples. However, only 722 frames have associated a LU. From those, only 9,360 LUs[3] where recognized by WordNet (out of 92%) corresponding to only 708 frames.

LUs of a frame can be nouns, verbs, adjectives and adverbs representing a coherent and closely related set of meanings that can be viewed as a small semantic field. For example, the frame EDUCATION_TEACHING contains LUs referring to the teaching activity and their participants. It is evoked by LUs like *student.n*, *teacher.n*, *learn.v*, *instruct.v*, *study.v*, etc. The frame also defines core semantic roles (or FEs) such as STUDENT, SUBJECT or TEACHER that are semantic participants of the frame and their corresponding LUs (see example below).

[Bernard Lansky]$_{STUDENT}$ *studied* [the piano]$_{SUBJECT}$ [with Peter Wallfisch]$_{TEACHER}$.

The paper is organized as follows. After this short introduction, in section 2 we present the graph-based Word Sense Disambiguation algorithm and the additional versions studied in this work. The evaluation framework and the results obtained by the different algorithms are presented and analyzed in section 3. Section 4 shows some examples and figures of the resulting semantic resource, and finally, in section 5, we draw some final conclusions and outline future work.

2 SSI algorithms

We have used a version of the Structural Semantic Interconnections algorithm (SSI) called SSI-Dijkstra(Cuadros and Rigau, 2008)(Laparra and Rigau, 2009). SSI is a knowledge-based iterative approach to Word Sense Disambiguation (Navigli and Velardi, 2005). The original SSI algorithm is very simple and consists of an initialization step and a set of iterative steps.

Given W, an ordered list of words to be disambiguated, the SSI algorithm performs as follows. During the initialization step, all monosemous words are included into the set I of already interpreted words, and the polysemous words are included in P (all of them pending to be disambiguated). At each step, the set I is used to disambiguate one word of P, selecting the word sense which is closer to the set I of already disambiguated words. Once a sense is disambiguated, the word sense is removed from P and included into I. The algorithm finishes when no more pending words remain in P.

In order to measure the proximity of one synset (of the word to be disambiguated at each step) to a set of synsets (those word senses already interpreted in I), the original SSI uses an in-house knowledge base derived semi-automatically which integrates a variety of online resources (Navigli, 2005). This very rich knowledge-base is used to calculate graph distances between synsets. In order to avoid the exponential explosion of possibilities, not all paths are considered. They used a context-free grammar of relations trained on Sem-

[2] `http://framenet.icsi.berkeley.edu/`

[3] Word-frame pairs

Cor to filter-out inappropriate paths and to provide weights to the appropriate paths.

Instead, SSI-Dijkstra uses the Dijkstra algorithm to obtain the shortest path distance between a node and some nodes of the whole graph. The Dijkstra algorithm is a greedy algorithm that computes the shortest path distance between one node an the rest of nodes of a graph. BoostGraph[4] library can be used to compute very efficiently the shortest distance between any two given nodes on very large graphs. As (Cuadros and Rigau, 2008), we also use already available knowledge resources to build a very large connected graph with 99,635 nodes (synsets) and 636,077 edges (the set of direct relations between synsets gathered from WordNet[5] (Fellbaum, 1998) and eXtended WordNet[6] (Mihalcea and Moldovan, 2001). For building this graph we used WordNet version 1.6 and the semantic relations appearing between synsets and disambiguated glosses of WordNet 1.7. To map the relations appearing in eXtended WordNet to WordNet version 1.6 we used the automatic WordNet Mappings[7] (Daudé et al., 2003). On that graph, SSI-Dijkstra computes several times the Dijkstra algorithm.

Previously, the SSI-Dijkstra algorithm have been used for constructing KnowNets (Cuadros and Rigau, 2008).

Initially, the list I of interpreted words should include the senses of the monosemous words in W, or a fixed set of word senses. Note that when disambiguating a Topic Signature associated to a particular synset, the list I always includes since the beginning at least the sense of the Topic Signature (in our example *pupil#n#1*) and the rest of monosemous words of W. However, many frames only group polysemous LUs. In fact, a total of 190 frames (out of 26%) only have polysemous LUs. Thus, SSI-Dijkstra provides no results when there are no monosemous terms in W. In this case, before applying SSI, the set of the LUs corresponding to a frame (the words included in W) have been ordered by polysemy degree. That is, the less polysemous words in W are processed first.

Obviously, if no monosemous words are found, we can adapt the SSI algorithm to make an initial guess based on the most probable sense of the less ambiguous word of W. For this reason we implemented two different versions of the basic SSI-Dijkstra algorithm: **SSI-Dijkstra-FirstSenses-I** (hereinafter FSI) and **SSI-Dijkstra-AllSenses-I** (hereinafter ASI). Thus, these two versions perform as SSI-Dijkstra when W contains monosemous terms, but differently when W contains only polysemous words. In fact, FSI and ASI always provide an interpretation of W.

While FSI includes in I the sense having minimal cumulated distance to the first senses of the rest of words in W, ASI includes in I the sense having minimal cumulated distance to the all the senses of the rest of words in W. The rationale behind the FSI algorithm is that the most frequent sense for a word, according to the WN sense ranking is very competitive in WSD tasks, and it is extremely hard to improve upon even slightly (McCarthy et al., 2004). Thus, this algorithm expects that the first sense in WordNet will be correct for most of the words in W. Regarding ASI, this algorithm expects that the words in W (corresponding to a very close semantic field) will establish many close path connections between different synsets of the same word (because of the fine-grained sense distinction of WordNet).

At each step, both the original SSI and also the SSI-Dijkstra algorithms only consider the set I of already interpreted words to disambiguate the next word of P. That is, the remaining words of P are not used in the disambiguation process. In fact, the words in P are still not disambiguated and can introduce noise in the process. However, the knowledge remaining in P can also help the process. In order to test the contribution of the remaining words in P in the disambiguation process, we also developed two more versions of the basic SSI-Dijkstra algorithm. **SSI-Dijkstra-FirstSenses-P** (hereinafter FSP) and **SSI-Dijkstra-AllSenses-P** (hereinafter ASP). When a word is being disambiguated, these two versions consider the set I of already interpreted words of W and also the rest of words remaining in P. That is, at each step, the algorithm selects the word sense which is closer to the set I of already disambiguated words and the remaining words of P all together. While FSP selects the sense having minimal cumulated distance to I and the first senses of the words in P, ASP selects the sense having minimal cumulated distance to I and all the senses of the words in P.

[4]http://www.boost.org/doc/libs/1_35_0/libs/graph/doc/index.html

[5]http://wordnet.princeton.edu

[6]http://xwn.hlt.utdallas.edu

[7]http://www.lsi.upc.es/~nlp/tools/mapping.html

	FN	GS	10	VM
#Frames	708	372	195	299
Nouns	5.87	7.90	13.58	4.18
Verbs	5.77	6.49	9.70	11.32
Adjectives	2.49	3.24	5.36	1.27
Other	0.11	0.14	0.24	0.05
Not in WN	1.07	1.30	2.13	1.02
Monosemous	4.40	5.79	9.87	4.50
Polysemous	8.77	10.68	16.88	11.30
#senses	3.64	3.45	3.63	4.31
Total	14.24	17.77	28.88	16.82

Table 1: Number of frames and average distribution of words per frame of the different datasets

3 Experiments

We have evaluated the performance of the different versions of the SSI algorithm using the same data set used by (Tonelli and Pianta, 2009). This data set consists of a total of 372 LUs corresponding to 372 different frames from FrameNet1.3 (one LU per frame). Each LUs have been manually annotated with the corresponding WordNet 1.6 synset. This Gold Standard includes 9 frames (5 verbs and 4 nouns) with only one LU (the one that has been sense annotated). Obviously, for these cases, our approach will produce no results since no context words can be used to help the disambiguation process[8]. Table 1 presents the main characteristics of the datasets we used in this work. In this table, *FN* stands for FrameNet[9], *GS* for the Gold-Standard, *10* for those Gold-Standard frames having at least 10 LUs and *VM* for the FrameNet–WordNet verb sense mapping[10] from (Shi and Mihalcea, 2005). Note that *VM* refers here to the characteristics of the frames appearing in the resource, not the mapping itself. The table shows for each dataset, the number of frames and the average distribution per frame of each POS, the words not represented in WordNet, the number of monosemous and polysemous words, the polysemy degree and the total words. The number of words per frame in this Gold Standard seems to be higher than the average in FrameNet.

Table 2 presents detailed results per Part-of-Speech (POS) of the performance of the different SSI algorithms on the Gold Standard in terms of Precision (P), Recall (R) and F1 measure (harmonic mean of recall and precision). In bold appear the best results for precision, recall and F1 measures. As baseline, we also include the performance measured on this data set of the most frequent sense according to the WordNet sense ranking (*wn-mfs*). Remember that this baseline is very competitive in WSD tasks, and it is extremely hard to beat.

We also included in the empirical evaluation the WordNet–FrameNet Verbal Mapping (**VM**) from (Shi and Mihalcea, 2005). As they work focused on verbal predicates for Semantic Parsing, *VM* does not provide provides results for nouns and adjectives. The annotation process between FrameNet 1.2 verb LUs and WordNet 2.0 verbal senses was performed manually. Since WordNet sense distinctions are very fine-grained, many verbal LUs in FrameNet were associated to multiple WordNet senses. We transport WordNet 2.0 sensekeys to version 1.6 by using the sense mappings from WordNet. Obviously, both FrameNet versions, that is 1.2 and 1.3, also presents differences in coverage of frames, LUs and correspondences between them. The final mapping covers a total of 299 frames and 2,967 verbal LUs.

As expected, the manual annotation provided by *VM* obtains the best results for verbs. However, possibly because of the different coverage of the FrameNet and WordNet versions, the recall is not as high as expected. In fact, the best recall for verbs is obtained by FSP.

All the different versions of the SSI-Dijkstra algorithm outperform the baseline. Only SSI-Dijkstra obtains lower recall for verbs because of its lower coverage. In fact, SSI-Dijkstra only provide answers for those frames having monosemous LUs, the SSI-Dijkstra variants provide answers for frames having at least two LUs (monosemous or polysemous) and the baseline always provides an answer.

As expected, the SSI algorithms present different performances according to the different POS. Also as expected, verbs seem to be more difficult than nouns and adjectives as reflected by both the results of the baseline and the SSI-Dijkstra algorithms. For nouns and adjectives, the best results are achieved by both FSI and ASI variants. Remember that these versions perform as SSI-Dijkstra on frames having monosemous LUs but performing an initial guess on frames having only polysemous LUs. While FSI makes an ini-

[8] In fact, FrameNet has 33 frames with only one LU, and 63 with only two.

[9] We removed frames without assigned LUs or not represented in WordNet

[10] Available at `http://www.cse.unt.edu/~rada/downloads.html`

	nouns			verbs			adjectives			all		
	P	**R**	**F**	**P**	**R**	**F**	**P**	**R**	**F**	**P**	**R**	**F**
VM	0.00	0.00	0.00	0.93	0.66	0.77	0.00	0.00	0.00	0.93	0.34	0.50
wn-mfs	0.75	0.75	0.75	0.64	0.64	0.64	0.80	0.80	0.80	0.69	0.69	0.69
SSI-Dijkstra	**0.84**	0.65	0.73	0.70	0.56	0.62	**0.90**	0.82	0.86	**0.78**	0.63	0.69
FSI	0.80	**0.77**	**0.79**	0.66	0.65	0.65	0.89	**0.89**	**0.89**	0.74	0.73	0.73
ASI	0,80	**0,77**	**0,79**	0,67	0,65	0,66	0,89	**0,89**	**0,89**	0,75	0,73	0,74
FSP	0.75	0.73	0.74	**0.71**	**0.69**	**0.70**	0.79	0.79	0.79	0.73	0.72	0.72
ASP	0.72	0.69	0.70	0.68	0.66	0.67	0.75	0.75	0.75	0.70	0.69	0.69
SSI-Dijkstra+	0.79	0.77	0.78	0.70	0.68	0.69	0.89	**0.89**	**0.89**	0.76	**0.74**	**0.75**

Table 2: Results of the different SSI algorithms on the *GS* dataset

tial guess including in I the sense of the less polysemous word having minimal cumulated distance to the *first senses* of the rest of words in W, ASI makes an initial guess including in I the sense of the less polysemous word having minimal cumulated distance to *all* the senses of the rest of words in W. In fact, FSI and ASI behave differently than SSI-Dijsktra in 73 frames having only polysemous LUs in the data set. Interestingly, the best results for verbs are achieved by FSP, not only on terms of F1 but also on precision. Remember that FSP always uses I and the first senses of the rest of words in P as context for the disambiguation. It seems that for verbs it is useful to consider not only the disambiguated words but also the most frequent senses of the rest of words being disambiguated. However, for nouns and adjectives the best precision is achieved by the original SSI-Dijkstra. This fact suggests the importance of having monosemous or correctly disambiguated words in I at the beginning of the incremental disambiguation process, at least for nouns and adjectives.

To our knowledge, on the same dataset, the best results so far are the ones presented by (Tonelli and Pianta, 2009) reporting a Precision of 0.71, a Recall of 0.62 and an F measure of 0.66[11]. Although they present a system which considers the most frequent sense, the most frequent domain and overlappings between the WordNet glosses and the FrameNet definitions of the LUs, in fact these results are below the most-frequent sense according to the WordNet sense ranking.

As a result of this empirical study, we developed **SSI-Dijkstra+** a new version of SSI-Dijkstra using ASI for nouns and adjectives, and FSP for verbs. SSI-Dijkstra+ clearly outperforms the baseline. Interestingly, the performance of this new algorithm improves overall, but obtains lower results for nouns than FSI and ASI and lower results for verbs than FSP.

[11]In fact, both evaluations are slightly different since they divided the dataset into a development set of 100 LUs and a testset with the rest of LUs, while we provide results for the whole dataset.

	P	R	F
mfs-wn	0.67	0.67	0.67
SSI-Dijkstra	**0.79**	0.74	0.76
FSI	0.78	0.78	0.78
ASI	0.78	0.77	0.78
FSP	0.72	0.71	0.71
ASP	0.70	0.70	0.70
SSI-Dijkstra+	**0.79**	**0.79**	**0.79**

Table 3: Results using FrameNet–WordNet Verbal mapping from (Shi and Mihalcea, 2005) as gold standard

Table 3 shows presents detailed results of the performance of the different SSI algorithms on the FrameNet–WordNet Verbal mapping (*VM*) produced by (Shi and Mihalcea, 2005) in terms of Precision (P), Recall (R) and F1 measure. In bold appear the best results for precision, recall and F1 measures. Again, we also include on the most frequent sense according to the WordNet sense ranking (*wn-mfs*).

On this dataset, the overall results are much higher because this dataset provides several correct verbal senses per LU. Again, the knowledgabased WSD algorithms perform over the most frequent sense baseline.

In fact, we expect much better results performing the disambiguation process including in I, when available, the manually assigned FrameNet–WordNet Verbal mappings. Possibly, using this approach very high accuracies for nouns, adjectives and the remaining verbs could be obtained.

4 WordFrameNet

The contribution of this new resource we call WordFrameNet is threefold[12]. First, we extend

[12]Available at `http://anonymous-web-page`

the coverage of FrameNet. That is, by establishing synset mappings to the FrameNet LUs, we can also add their corresponding synonyms to the frame. For instance, the frame EDUCATION_TEACHING only considers *instruct.v* and *instruction.n*, but not *instructor.n* which is a synonym in WordNet of the LU *teacher.n*. Thus, while the original FrameNet have 9,328 LUs corresponding to 6,565 synsets, WordFrameNet have 20,587 LUs. That is, more than the double. Tables 5 and 6 show respectively, the original and new LUs for the EDUCATION_TEACHING frame. In this case, 24 of the original LUs have been associated to WN synsets, thus producing 18 new LUs for this frame.

Second, we can extend the coverage of semantic relations in WordNet. That is, by establishing new semantic relations among all the LUs of a particular frame. For instance, in WordNet there is no a semantic relation connecting <*student*, *pupil*, *educatee*> and < *teacher*, *instructor*> directly. In that way, we obtain 124,718 new semantic relations between the original 6,565 synsets. 121,813 of these relations that connect synsets of the same frame do not appear in WordNet. In table 4 we show the number of existing WordNet relations between synsets of the same frame.

Hypernymy	2028
Antonymy	408
Similar-to	328
Also-see	178
Part	97
Attribute	82
Entailment	44
Verb-group	22
Derived-from	18
Cause	14
Member	8
Substance	4
Participle-of-verb	1

Table 4: WordNet relations in FrameNet

Third, we can also automatically extend FrameNet to languages other than English by exploiting local wordnets aligned to the English WordNet. For instance, the Spanish synset aligned to <*student*, *pupil*, *educatee*> is <*alumno*, *estudiante*> and the Italian one is <*allievo*, *alunno*, *studente*>. In Spanish, we obtain a WordFrameNet with 14,106 LUs. In fact, the current version of the Spanish FrameNet consists of 308 frames with 1,047 LUs[13] (Subirats and Petruck, 2003). Table 7 presents the Spanish version of WordFrameNet for the EDUCATION_TEACHING frame. In this case, 30 Spanish LUs have been associated to this particular frame, while the current version of the Spanish FrameNet only have 2 LUs (*aprender.v* and *enseñar.v*).

Furthermore, we can also transport to the disambiguated LUs the knowledge currently available from other semantic resources associated to WordNet such as SUMO (Niles and Pease, 2001), WordNet Domains (Magnini and Cavaglià, 2000), etc. For instance, now the LU corresponding to *student.n* can also have associated the SUMO label *SocialRole* and its corresponding logical axioms, and the WordNet Domains *school* and *university*.

5 Conclusions and future work

In this work, we have presented an ongoing work aiming to integrate FrameNet and WordNet. The method uses a knowledge based Word Sense Disambiguation (WSD) algorithm called SSI-Dijkstra for assigning the appropriate synset of WordNet to the semantically related Lexical Units of a given frame from FrameNet. This algorithm relies on the use of a large knowledge base derived from WordNet and eXtended WordNet. Since the original SSI-Dijkstra requires a set of monosemous or already interpreted words, we have devised, developed and empirically tested different versions of this algorithm to deal with sets having only polysemous words. The resulting new algorithms obtain improved results over state-of-the-art.

Finally, using the same automatic approach, we also plan to disambiguate the Lexical Elements of a given frame. Thus, the resulting resource will also integrate the core semantic roles of FrameNet. For example, for the frame EDUCATION_TEACHING we will associate the appropriate WordNet synsets to the Lexical elements STUDENT, SUBJECT or TEACHER.

[13] `http://gemini.uab.es:9080/SFNsite/sfn-data/current-project-status`

train.v	00407541-v	prepare for a future task or career
instruct.v	00562446-v	impart skills or knowledge to
educational.a	02716766-a	relating to the process of education
tutee.n	07654181-n	learns from a tutor
schoolteacher.n	07551404-n	a teacher in a school below the college level
educate.v	00407541-v	prepare for a future task or career
study.v	00405251-v	be a student of a certain subject
instruction.n	00567704-n	activities that impart knowledge
teacher.n	07632177-n	a person whose occupation is teaching
student.n	07617015-n	a learner who is enrolled in an educational institution
schoolmistress.n	07550942-n	a woman schoolteacher
tutor.v	00562981-v	be a tutor to someone; give individual instruction
lecturer.n	07367816-n	someone who lectures professionally
training.n	00574678-n	activity leading to skilled behavior
pupil.n	07617015-n	a learner who is enrolled in an educational institution
schoolmaster.n	07551048-n	any person (or institution) who acts as an educator
school.v	01626656-v	educate in or as if in a school
master.v	00403563-v	be or become completely proficient or skilled in
tutor.n	07162304-n	a person who gives private instruction (as in singing or acting)
professor.n	07504465-n	someone who is a member of the faculty at a college or university
learn.v	00562446-v	impart skills or knowledge to
teach.v	00562446-v	impart skills or knowledge to
coach.v	00565367-v	teach and supervise, as in sports or acting
education.n	00567704-n	activities that impart knowledge

Table 5: LUs corresponding to EDUCATION_TEACHING frame

develop.v	00407541-v
prepare.v	00407541-v
educate.v	00407541-v
instruct.v	00562446-v
school_teacher.n	07551404-n
read.v	00405251-v
take.v	00405251-v
teaching.n	00567704-n
pedagogy.n	00567704-n
educational_activity.n	00567704-n
instructor.n	07632177-n
educatee.n	07617015-n
schoolmarm.n	07550942-n
mistress.n	07550942-n
preparation.n	00574678-n
grooming.n	00574678-n
get_the_hang.v	00403563-v
private_instructor.n	07162304-n

Table 6: New LUs associated to EDUCATION_TEACHING frame

adiestrar.v	00407541-v
amaestrar.v	00407541-v
enseñar.v	00562446-v
instruir.v	00562446-v
educacional.a	02716766-a
maestra.n	07551404-n
maestro.n	07551404-n
profesor.n	07551404-n
profesora.n	07551404-n
aprender.v	00405251-v
educación.n	00567704-n
enseñanza.n	00567704-n
instructor.n	07632177-n
monitor.n	07632177-n
profesor.n	07632177-n
alumna.n	07617015-n
alumno.n	07617015-n
estudiante.n	07617015-n
maestra.n	07550942-n
profesora.n	07550942-n
tutelar.v	00562981-v
conferenciante.n	07367816-n
formación.n	00574678-n
preparación.n	00574678-n
instructor.n	07162304-n
preceptor.n	07162304-n
profesor_particular.n	07162304-n
tutor.n	07162304-n
profesor.n	07504465-n
entrenar.v	00565367-v

Table 7: Spanish LUs inferred for EDUCATION_TEACHING frame

References

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

J. Atserias, L. Villarejo, G. Rigau, E. Agirre, J. Carroll, B. Magnini, and Piek Vossen. 2004. The meaning multilingual central repository. In *Proceedings of GWC*, Brno, Czech Republic.

C. Baker, C. Fillmore, and J. Lowe. 1998. The berkeley framenet project. In *COLING/ACL'98*, Montreal, Canada.

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.

M. Cuadros and G. Rigau. 2008. Knownet: Building a large net of knowledge from the web. In *22nd International Conference on Computational Linguistics (COLING'08)*, Manchester, UK.

J. Daudé, L. Padró, and G. Rigau. 2003. Validation and Tuning of Wordnet Mapping Techniques. In *Proceedings of RANLP*, Borovets, Bulgaria.

Katrin Erk and Sebastian Pado. 2004. A powerful and versatile xml format for representing role-semantic annotation. In *Proceedings of LREC-2004*, Lisbon.

C. 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.

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.

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

Egoitz Laparra and German Rigau. 2009. Integrating wordnet and framenet using a knowledge-based word sense disambiguation algorithm. In *Proceedings of RANLP*, Borovets, Bulgaria.

B. Magnini and G. Cavaglià. 2000. Integrating subject field codes into wordnet. In *Proceedings of LREC*, Athens. Greece.

D. McCarthy, R. Koeling, J. Weeds, and J. Carroll. 2004. Finding predominant senses in untagged text. In *Proceedings of ACL*, pages 280–297.

R. Mihalcea and D. Moldovan. 2001. extended wordnet: Progress report. In *Proceedings of NAACL Workshop on WordNet and Other Lexical Resources*, Pittsburgh, PA.

R. Navigli and P. Velardi. 2005. Structural semantic interconnections: a knowledge-based approach to word sense disambiguation. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 27(7):1063–1074.

R. Navigli. 2005. Semi-automatic extension of large-scale linguistic knowledge bases. In *Proc. of 18th FLAIRS International Conference (FLAIRS)*, Clearwater Beach, Florida.

I. Niles and A. Pease. 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.

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

Marco Pennacchiotti, Diego De Cao, Roberto Basili, Danilo Croce, and Michael Roth. 2008. Automatic induction of FrameNet lexical units. In *Proceedings of Empirical Methods in Natural Language Processing (EMNLP)*, Honolulu, Hawaii, USA.

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 CICLing*, Mexico.

Carlos Subirats and Miriam R.L. Petruck. 2003. Surprise: Spanish framenet! In *Proceedings of the International Congress of Linguists*, Praga.

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.