

Extensive Evaluation of a FrameNet-WordNet mapping resource

Diego De Cao, Danilo Croce, Roberto Basili

University of Roma Tor Vergata, Roma, Italy
{decao,croce,basili}@info.uniroma2.it

Abstract

Lexical resources are basic components of many text processing system devoted to information extraction, question answering or dialogue. In past years many resources have been developed such as FrameNet and WordNet. FrameNet describes prototypical situations (i.e. Frames) while WordNet defines lexical meaning (senses) for the majority of English nouns, verbs, adjectives and adverbs. A major difference between FrameNet and WordNet refers to their *coverage*. Due of this lack of coverage, in recent years some approaches have been studied to make a bridge between this two resources, so a resource is used to extend the coverage of the other one. The nature of these approaches leave from supervised to supervised methods. The major problem is that there is not a standard in evaluation of the mapping. Each different work have tested own approach with a custom gold standard. This work give an extensive evaluation of the model proposed in (De Cao et al., 2008) using gold standard proposed in other works. Moreover this work give an empirical comparison between other available resources. As outcome of this work we also release the full mapping resource made according to the model proposed in (De Cao et al., 2008).

1. Introduction

Lexical resources are basic components of many language processing system devoted to information extraction, question answering or dialogue. Several approaches to lexical semantics, such as wordnets or frame semantic dictionaries, gave rise to large scale resources, respectively WordNet (Miller et al., 1990) or FrameNet (Baker et al., 1998). FrameNet describes prototypical situations (i.e. Frames) through a number of associated frame-evoking words, the so-called *lexical units* (LU). Moreover, for each Frame a set of prototypical semantic arguments, called Frame Elements (FE), characterize all the participants to the underlying event. On the contrary, WordNet defines senses for the majority of English nouns, verbs, adjectives and adverbs in terms of *sets of synonyms*, called *synsets*. Each *synset* represents a lexical meaning.

A major difference between FrameNet and WordNet refers to their *coverage*. WordNet's size increased along the years, and the current version includes about 207.000 senses for about 155,000 different words/lemmas. FrameNet defines about 10,000 lexical units for about 800 frames. Recent studies (Shen and Lapata, 2007), while showing that the use of FrameNet is potentially beneficial in Question Answering systems, also point out that due to the low coverage of the current FrameNet the expected boost in performance is inherently limited. Other studies have shown similar evidences for Recognizing Textual Entailment (RTE) (Clark et al., 2007; Burchardt et al., 2008): most examples of the RTE challenges corpora can be solved at the predicate-argument structure level, but FrameNet coverage is still a major problem.

Possibilities to extend the lexical coverage is to automatically or semi-automatically acquire lexical units from corpora as explored in (Pennacchiotti et al., 2008). Alternatively, one could extend the FrameNet coverage by making a bridge with larger resources, such as WordNet (or VerbNet). Among other works concerning the mapping between lexical units and synsets, (Burchardt et al., 2005) discusses Detour, a system for predicting frame assignment of potential lexical units not covered by FrameNet. Detour is

based on paradigmatic information enclosed in WordNet (i.e. hyponymy relationships between word senses). Although the authors do not fully solve the underlying disambiguation problem between senses and frames, they propose an empirical association measure that ranks frame candidates for each sense as defined in WordNet. In (Shi and Mihalcea, 2005) a model to automatic map FrameNet verbal lexical units to VerbNet verbs (Levin, 1993; Kipper et al., 2000) using WordNet as a bridge is also presented. (Pitel, 2006) presents a preliminary study on the applicability of semantic spaces and space geometrical transformations (namely, Latent Semantic Analysis) to expand FrameNet, but the investigation is too limited in scope to draw relevant conclusions.

Recently, in (Tonelli and Pighin, 2009) a mapping between FrameNet Lexical Units and WordNet synsets is studied as a classification task according to a supervised learning model. An SVM is employed to decide if a candidate WordNet synset corresponds to a specific Lexical Unit of a Frame. The main information used by the classifier is made available by the glosses of the frame definitions, of the Wordnet senses and of the individual lexical units.

In (De Cao et al., 2008), we proposed an unsupervised model for inducing Lexical Units by combining distributional, i.e. corpus, evidence as well as paradigmatic information derived from Wordnet. As a side effect of that work, a mapping of Lexical Units to WordNet synsets is derived. Moreover, as discussed in (De Cao et al., 2008; Tonelli and Pighin, 2009), the mapping allows easily to extend FrameNet to a foreign language, through parallel WordNet resources, such as MultiWordNet (Pianta et al., 2002) for the Italian language. This can lay the foundation to the development of a full FrameNet for Italian as reported in (De Cao et al., 2008).

Hereafter, we summarize the paradigmatic model introduced in (De Cao et al., 2008) in section 2. while section 3. describe and evaluate the obtained resource. An extended evaluation with comparison between the other available resources reported in (Padó et al., 2008) and (Shi and Mihalcea, 2005) will be reported in final version of this work.

2. Using Paradigmatic information for mapping senses to frames

The basic intuition behind the paradigmatic model presented in (De Cao et al., 2008) is that the knowledge of the entire set of lexical units of a given frame allows to better model the intended meaning of an individual lexical unit and select the proper subset of its WordNet senses: these are the only suitable to evoke the underlying frame and the other are neglected. In (De Cao et al., 2008), we thus assume that these senses are topologically related to (one or more) WordNet sub-hierarchies capturing the lexical semantics implicit in the frame. So, frames correspond to specific sub-graphs of the WordNet hyponymy hierarchy. Figure 1 reports the WordNet sub-hierarchy covering the frame PEOPLE_BY_AGE: here, the frame’s nominal LUs {*adult*, *adolescent*, *baby*, *boy*, *infant*, *kid*, *geezer*, *teenager*, *youngster*, *youth*} are all represented with the senses correctly referring to the frame. The correct senses (e.g. sense 1 of *youth* out of its 6 potential senses) are selected as they share most specific generalizations with the other LUs. This graph can be intended as an “*explanation*” of the lexical semantic properties characterizing the frame. We call such a graph the *Paradigmatic* model of the frame. As WordNet organizes nouns, verbs and other parts-of-speech in different hierarchies, three independent *Paradigmatic* models (one for each part-of-speech) are created for each frame.

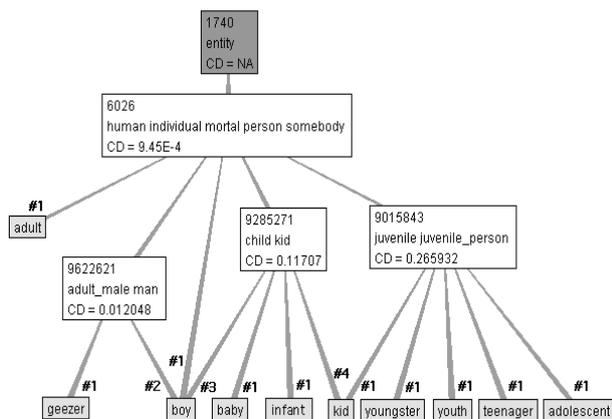


Figure 1: The WordNet model for the frame *People.by.Age* as evoked by the set of its nouns. Sense numbers #*n* refer to WordNet 2.0.

Models for each frame were built: they give rise to a score for each *lexical sense* of a LU based on a similarity function sim_{WN} , that is independently defined for verbs, nouns and adjectives.

For **nouns**, we adopt the *conceptual density* (*cd*) (Agirre and Rigau, 1996; Basili et al., 2004), a semantic similarity measure defined for word sense disambiguation tasks. The *cd* score for a sense σ of a noun is the density of the WordNet sub-hierarchy rooted at σ in representing the entire set of all nouns in the frame F ¹. The intuition behind this model is that the larger is the number of all and only nominal LUs in F that are generalized by a synset σ , the better σ

¹In the following, we will use the same notation for a frame F and for the set of its known lexical units.

captures the lexical semantics intended by the frame. Figure 1 shows that correct senses (e.g. the sense 1 of *youth* out of the 6 potential senses) are generally detected and preserved in the model. Irrelevant senses that do not share any common hypernym with other words in F are neglected. So the *cd* score is used to rank the individual senses as in the case of *boy* in Figure 1. The $sim_{WN}(F, n)$ for the noun n , given a Frame F , is given by the highest *cd* score across all its lexical senses σ .

As conceptual density can be only applied to nouns, when **verbs** v are considered, we exploit the synonymy and co-hyponymy relations. The following similarity $sim_{WN}(F, v)$ is computed:

$$sim_{WN}(F, v) = \begin{cases} 1 & \text{iff } \exists K \subset F \text{ such that } |K| > \tau \text{ AND} \\ & \forall w \in K \quad w \text{ is a co-hyponym of } v \\ \epsilon & \text{otherwise} \end{cases} \quad (1)$$

For **adjectives**, the similarity $sim_{WN}(F, a)$, is computed on the basis of the synonymy relation, as follows:

$$sim_{WN}(F, a) = \begin{cases} 1 & \text{iff } \exists w \in F \text{ such that} \\ & w \text{ is a synonym of } a \\ \epsilon & \text{otherwise} \end{cases} \quad (2)$$

Table 3 shows an example of the output of the paradigmatic model for four words, i.e. *room*, as a noun, to *chain*. Scores here are probability estimates $p(F|w, \sigma)$ derived from the definitions in Eq.2 and 1: in particular for the verb *rotate.v* the score 0.6 tells us that 60% of the co-hyponyms found in CAUSE_TO_MOVE_IN_PLACE (i.e. that form the set K in Eq. 1) are justified by that sense (i.e. “*turn on or around an axis or a center*”).

3. Empirical Analysis

The method to map a FrameNet lexical units into one or more WordNet synsets, presented in (De Cao et al., 2008) and summarized in Section 2., was able to propose about 10,000 instances of <Frame, LU, Synset> triples, that involve 619 frames. The mapping is defined over the FrameNet 2.0 and WordNet 2.0. Table 1 reports the overall statistics for the resource described in the (De Cao et al., 2008) paper. The 3,602 nouns, 3,325 verbs and 762 adjectives examined gave rise to a large overall ambiguity, as the number of involved senses (i.e. the candidate lexical senses suggested by at least one lexical unit in the fourth row) suggests. The result is an average polysemy between 3 and 6 Wordnet senses per predicate (fifth row). The adoption of the paradigmatic models results in a significant reduction in the average ambiguity: only 1.46 senses per verb survive among the initial 5.64, while about 1.2 among the 3 senses are retained for nouns and adjectives.

In order to evaluate the accuracy of these models, we carried out in (De Cao et al., 2008) an evaluation of four automatically generated frames, against a manually build gold standard. This showed a precision of 0.803 and a recall of 0.79 (F-measure=0.796). However, the analysis was limited to a small set of lexical units and no general assess of our approach was possible. Due to the complexity of the

	Nouns	Verbs	Adjectives
Targeted Frames	364	412	111
Targeted LUs	3.602	3.325	762
Average LUs per frame	9,89	8,07	6,86
Number of Evoked Senses	11.034	18.781	2.320
Average Polysemy	3,06	5,64	3,04
Active Lexical Senses	4.221	4.868	921
Average Active Lexical Senses per word over frames	1,17	1,46	1,20

Table 1: Statistics on nominal, verbal and adjectival senses in the paradigmatic model of the English FrameNet

mapping problem, a comprehensive analysis for these researches needs more representative resources.

In this paper, we increase the size of our analysis by considering also the same gold standard of (Tonelli and Pighin, 2009). Although the *paradigmatic* models suggested in (De Cao et al., 2008) were trained on WordNet 2.0, the oracle used in (Tonelli and Pighin, 2009) is based on WordNet 1.6. A correspondence between two different version was made using the mappings made available by Rada Mihalcea². Hereafter the oracle of (Tonelli and Pighin, 2009) mapped into its Wordnet 2.0 version will be referred as *TP*. The *TP* oracle includes 386 Frames with a set of 2,158 manually validated LU-synset pairs. As our paradigmatic model uses different metrics for different Part-of-speeches, in Table 2 results are also independently reported for nouns, verbs and adjectives. There, the results achieved by two different polynomial kernels, as discussed in (Tonelli and Pighin, 2009), are also shown. This initial investigation suggests that our method performs much better for nouns and adjectives, as the corresponding metrics seem to better govern the underlying polysemy effects. Verbs are a much more complex category, whose polysemy is much higher on average, i.e. 5.64 senses per lexical unit. The simpler metric based only on the co-hyponymy relation, as defined by Eq. 1, seems to achieve a much lower accuracy. This reflects well known problems related to the disambiguation of verb senses (a much more complex problem) as well as the inherent topological differences between the noun and verb lexical hierarchies in Wordnet.

	Precision	Recall	F-Measure
Tonelli-Pighin 1	0,761	0,613	0,679
Tonelli-Pighin 2	0,794	0,569	0,663
Noun	0,795	0,815	0,805
Verb	0,522	0,665	0,585
Adjectives	0,694	0,735	0,714

Table 2: Results against the *TP* oracle, proposed in (Tonelli and Pighin, 2009), for different Part-of-speech.

In Table 3, some examples of the output of the *paradigmatic model* are reported. For each lexical unit, the paradigmatic similarity score (third column) and the number of senses in WordNet (fourth column) are shown. While for

the first two rows (i.e. *room.n* and *flow.v*), the selected senses are those confirmed by the manual oracle *TP* of (Tonelli and Pighin, 2009), the last two rows (i.e. *rotate.v* and *chain.n*) describe the selected sense is judged as not correct. Notice that the most appropriate sense of the LU *rotate.v* explicitly refers to *causality*, although it has a definition, i.e. *cause to turn on an axis or center*; “Rotate the handle” very close to the one proposed by our system (row fourth, last column).

4. Comparative Analysis

While the previous section discussed the performances as we can possibly measure against a gold standard, it must be noticed that the number of employed test instances is not very large if compared with the entire set of Lexical Unit currently available from FrameNet. Table 1 for example shows that the overall number of instances processed in (De Cao et al., 2008) is much larger than the adopted gold standard, the latter being limited to about the 8% of the overall set. At the same time, we also know that other efforts (such as the work discussed in (Tonelli and Pighin, 2009) and (Shi and Mihalcea, 2005)) made available other resources on a realistic scale. This allows to design a more comprehensive comparative analysis where one resource can be seen as the gold-standard for the other ones. This allow to triangulate across resources and to get a better picture of benefits or limitations of the different approaches. According to the above view, we can carry out two kinds of analysis. The first is centered on the quantitative evaluation of the agreement between all the involved resource pairs. In this way, we can proceed in the most neutral way across resources and gather objective measures of the effectiveness of the individual approaches. The second analysis is instead qualitative and concerns with the discussion of cases of disagreement between the underlying approaches.

4.1. Quantitative Analysis

In this analysis, we focus only on the lexical units that have been considered and somehow annotated in all the three methods:

- the paradigmatic *PM* model of (De Cao et al., 2008)
- the SVM-based method of (Tonelli and Pighin, 2009) hereafter *TP* and

²<http://www.cse.unt.edu/~rada/downloads.html>

- the Framenet to Wordnet maps of (Shi and Mihalcea, 2005), hereafter *F2W*)

The experiments below are thus focusing on all LU-frame pairs (w, F) for which one sense σ of w has been associated to w given F by all targeted methods. If only *TP* and *PM* are considered, we have 3,479 such (w, F) pairs. If we took into account all the three resources, *PM*, *TP* and *F2W*, then the number of common pairs is 1,027³.

As a metric, we adopted Cohen’s kappa. In general, the Cohen’s Kappa measure has been used to evaluate the agreement between manually annotated data. Our hypothesis is that every resource corresponds to a manual annotation (i.e. a gold standard) in a test, and the Cohen’s Kappa measures of other resources capture how much differently the methods label the same cases. Moreover the *Agreement* score expresses the percentage of the decisions (i.e. (w, σ, F) targeted LU, synset and frame triples) for which the decision of the two method was the same, i.e. a lexical sense σ of w is accepted or rejected for the pair (w, F) by both models. In Table 4 the Cohen’s Kappa factor between the *PM* model and the *TP* ((Tonelli and Pighin, 2009)) is reported. The *Agreement* scores are reported in the third column of Table 4. Notice that we still have disagreement for about two thousand (w, σ, F) triples, for about 1,148 words w (about the 33% of the target LUs). This suggests that the most ambiguous words are also those for which most errors are done.

	Cohen’s k	Agreement
Overall	0,69	86,0%
Noun	0,70	85,3%
Verb	0,65	86,7%
Adjectives	0,69	85,2%

Table 4: Results of Cohen’s Kappa statistics between the *PM* model and the *TP* model of (Tonelli and Pighin, 2009).

We also compared our model against the Frame-to-Wordnet maps proposed in (Shi and Mihalcea, 2005), where the verbal lexical units defined in FrameNet are mapped to their corresponding WordNet synsets. The Cohen’s Kappa results against this resource is reported in Table 5, second row. This comparison is restricted to verbal lexical units (the only ones treated in (Shi and Mihalcea, 2005)). For sake of comparison, we also report (first row in Table 5) the Cohen’s kappa statistics of our *PM* model against the *TP* resource (Tonelli and Pighin, 2009), over the same subset of lexical units (i.e. the verbs).

A general outcome is that, in every test, a large number of the involved predicates (or lexical units), e.g. the 68% of the overall set in Table 4, is classified in the same way by the two targeted methods. The k value suggests a substantial agreement about triples (w, σ, F) . If we focus just on the matching between frames F and synset σ , rather than also considering the individual words w , the agreement grows up: for example 86% of the possible senses

³Notice that in (Shi and Mihalcea, 2005) only verbs are considered

	Cohen’s k	Agreement
MapNet (<i>TP</i> verbs only)	0,65	85,8%
FnWnVerbMap (<i>F2W</i>)	0,58	82,5%

Table 5: Results of Cohen’s Kappa statistics of the *PM* model against the *TP* model in (Tonelli and Pighin, 2009) and the *F2W* model in (Shi and Mihalcea, 2005).

for a frame F are accepted (or rejected) by both *PM* and *TP* methods (Table 4), thus suggesting an almost perfect agreement. This outcome is rather interesting as it confirms that senses are correlated to predicates in general, although the two semantic notions are originally independent. As a consequence, employing the information about the former to get knowledge about the latter is effective. Notice that this confirms the soundness of the research works devoted to exploit sense information in the automatic induction of frame information (as explored in (Burchardt et al., 2005) or (Pennacchiotti et al., 2008)).

A second general consequence of the above measures is related to the non negligible subset of data where the systems disagree. For a not-so-small dataset the mappings proposed by two models are different. While this can be expected, this also suggests that the systems embody independent views of the source data. Notice how the model in (Tonelli and Pighin, 2009) strongly relies on the lexical information made available by the definitions in both Framenet and Wordnet, while, on the contrary, the paradigmatic model in (De Cao et al., 2008) exploits the topological structure of Wordnet to model the similarity between sets of lexical units (i.e. the entire frame) and a single predicate. The independence of the two methods suggest that further improvements on the two methods can be obtained by applying them in combination. The following qualitative analysis of the results of our tests seems to confirm this thesis.

4.2. Qualitative Analysis

The manual analysis of those mappings for which *PM* and *TP* suggest different annotations aims at capturing the nature and causes of the disagreement. Some examples are reported in Table 6. Notice how the first two rows in Table 6 refer to Frames, ACCOUTREMENTS and GROOMING, for which the two models suggest different but equally acceptable senses. For example the two senses of the verb *soap* selected respectively by *PM* and *TP* are both valid for GROOMING: they correspond to distinctions made at a finer grain than the one required by the frame itself. It is often the case, due to the fine grain of the employed definitions, several synsets are equally acceptable for a frame. The fact that the *PM* and *TP* models differ depend on the independent views they express about the resources, i.e. the lexical sense network topology in *PM* vs. the glosses for *TP*. On the opposite, in the third row only the *PM* sense is correct, while in the last row the reverse is true (i.e. *TP* provides the valid sense). In general, while in most cases both systems give the same (seemingly correct) answers, subtle distinctions can be found in disagreement. Notice

Frame	Frame Def.	Lexical Unit	Score	Senses	WordNet Gloss
BUILDING_SUBPARTS	This frame includes words that name subparts of buildings that can be occupied by people.	<i>room.n</i>	1	4	an area within a building enclosed by walls and floor and ceiling; “the rooms were very small but they had a nice view”
FLUIDIC_MOTION	In this frame a Fluid moves from a Source to a Goal along a Path or within an Area.	<i>flow.v</i>	0.9	7	move along, of liquids; “Water flowed into ; the cave” “the Missouri feeds into the Mississippi”
CAUSE_TO_MOVE_IN_PLACE	An Agent causes a Theme to move with respect to a certain Fixed_location, generally with a certain Periodicity, ...	<i>rotate.v</i>	0.6	7	turn on or around an axis or a center; “The Earth revolves around the Sun”; “The lamb roast rotates on a spit over the fire”
CONNECTORS	The Connector is an artifact created to affix a Connected_item or to bind onto a Fixed_location and is primarily so used.	<i>chain.n</i>	0.69	10	a necklace made by a stringing objects together; “a string of beads”; “a strand of pearls”;

Table 3: An example of the output of the paradigmatic models

how the example about the lexical unit *electrical.a* suggests that mistakes are made by the *TP* model when the definition is particularly generic. The gloss of *electrical.a* in Table 6 is very general and this is the major cause of the error made by the *TP* model on *electrical.a*. Apparently, the *PM*, that employs a distance based on the Wordnet synset hierarchy, is more robust with respect to these misleading cases. On the contrary, the *PM* modeling tend to be confused when fine grained senses of a lexical unit are involved and the difference between them is very small: this seems to be suggested by the example on *stance.n*, where the *PM* fails. In all these cases, strengths and weaknesses are complementary. A proper combination of the *PM* and the *TP* model seems quite promising given that it can be very effective as a way to balance against the weak assumptions of the two models.

5. Conclusions

In this paper an extensive evaluation of a FrameNet to WordNet mapping model is presented. The evaluation was carried out both against a manually annotated gold standard previously employed and through a cross-validation involving three different automatically acquired resources. The resource based on the paradigmatic model here proposed has been made publicly available at <http://sag.art.uniroma2.it/>. Its comparison with the gold standard already adopted in (Tonelli and Pighin, 2009) mainly confirms the previous accuracy measures reported in (De Cao et al., 2008). However, the cross-validation of the *TP* resource of (Tonelli and Pighin, 2009) with our paradigmatic model *PM* is interesting. The study of the two systems’ output suggest a substantial agreement between the two methods, that becomes almost perfect if

only synsets (i.e. not words) and frames are taken into account. As the two methods make use of independent information they can be effectively integrated within a structured supervised (e.g. SVM-based) approach, such as *stacking* or *late fusion*. This constitutes the core of future research work based on the outcomes of this paper. Moreover, all methods studied here are rather independent on the corpus analysis. Corpus information has been employed and shown useful in previous work (i.e. (De Cao et al., 2008)), instead. When the targeted Framenet to Wordnet mapping is tailored by a distributional analysis (i.e. through collocational lexical information) over the source corpus, the mapping can be even made dependent on the underlying domain, something that, on a linguistic basis, seems more appropriate for the complexity of most NLP applications.

Acknowledgements

We thank Sara Tonelli and Daniele Pighin, that made available the gold standard on the frame to synset mappings employed in this paper.

6. References

- E. Agirre and G. Rigau. 1996. Word sense disambiguation using conceptual density. In *Proceedings of COLING-96*, Copenhagen, Denmark.
- Collin F. Baker, Charles J. Fillmore, and John B. Lowe. 1998. The Berkeley FrameNet project. In *Proc. of COLING-ACL*, Montreal, Canada.
- R. Basili, M. Cammisa, and F.M. Zanzotto. 2004. A semantic similarity measure for unsupervised semantic disambiguation. In *Proceedings of LREC-04*, Lisbon, Portugal.

Frame	Frame Definition	LU	WordNet Gloss	System
ACCOUTREMENTS	A Wearer wears accessories, which are made of some Material and may have a Style.	<i>choker.n</i>	necklace that fits tightly around a woman's neck	PM
			a high tight collar	TP
GROOMING	In this frame, an Agent engages in personal body care. An Instrument can be used in this process as well as a Medium.	<i>soap.v</i>	rub soap all over, usually with the purpose of cleaning	PM
			cover with soap; "lather your body when you shower"	TP
ELECTRICITY	Lexical units in this frame refer to Electricity, in particular as a form of energy harnessed for particular uses (such as powering machines). The Source of the Electricity may also be expressed, or incorporated in the meaning of the LUs.	<i>electrical.a</i>	using or providing or producing or transmitting or operated by electricity; "electric current"; "electric wiring"	PM
			relating to or concerned with electricity; "an electrical engineer"; "electrical and mechanical engineering industries"	TP
POSTURE	An Agent supports their body in a particular Location. ...	<i>stance.n</i>	a rationalized mental attitude	PM
			standing posture	TP

Table 6: Examples of disagreement between the paradigmatic model *PM* and the *TP* model defined in (Tonelli and Pighin, 2009).

- Aljoscha Burchardt, Katrin Erk, and Anette Frank. 2005. A wordnet detour to framenet. In *Proceedings of the GLDV 2005 GermaNet II Workshop*, Bonn, Germany.
- Aljoscha Burchardt, Marco Pennacchiotti, Stefan Thater, and Manfred Pinkal. 2008. Assessing the impact of frame semantics on textual entailment. *Journal of Natural Language Engineering (to appear)*.
- Peter Clark, Phil Harrison, John Thompson, William Murray, Jerry Hobbs, and Christiane Fellbaum. 2007. On the role of lexical and world knowledge in rte3. In *Proceedings of the ACL-PASCAL Workshop on Textual Entailment and Paraphrasing*, pages 54–59, Prague, June. Association for Computational Linguistics.
- D. De Cao, Danilo Croce, Marco Pennacchiotti, and Roberto Basili. 2008. Combining word sense and usage for modeling frame semantics. In *Proceedings of STEP 2008*, Venice, Italy, September.
- Karin Kipper, Hoa Trang Dang, and Martha Palmer. 2000. Class-based construction of a verb lexicon. In *Proceedings of Seventeenth National Conference on Artificial Intelligence AAAI 2000*, Austin, TX, July.
- Beth Levin. 1993. English verb classes and alternations: A preliminary investigation. The University of Chicago Press.
- G. Miller, R. Beckwith, C. Fellbaum, D. Gross, and K. Miller. 1990. An on-line lexical database. *International Journal of Lexicography*, 13(4):235–312.
- Sebastian Padó, Marco Pennacchiotti, and Caroline Sporleder. 2008. Semantic role assignment for event nominalisations by leveraging verbal data. In *Proceedings of COLING 2008*, Manchester, UK.
- Marco Pennacchiotti, Diego De Cao, Roberto Basili, Danilo Croce, and Michael Roth. 2008. Automatic induction of framenet lexical units. In *EMNLP. ACL*.
- Emanuele Pianta, Luisa Bentivogli, and Christian Girardi. 2002. Multiwordnet: developing an aligned multilingual database". In *Proceedings of the First International Conference on Global WordNet*, Mysore, India, January 21-25.
- Guillaume Pitel. 2006. Using bilingual lsa for framenet annotation of french text from generic resources. In *Workshop on Multilingual Semantic Annotation: Theory and Applications*, Saarbrücken, Germany.
- Dan Shen and Mirella Lapata. 2007. Using semantic roles to improve question answering in proceedings emnlp-conll, 2007. In *Proceedings EMNLP-CoNLL*.
- Lei Shi and Rada Mihalcea. 2005. Putting pieces together: Combining framenet, verbnet and wordnet for robust semantic parsing. In *Cicling*, Mexico.
- Sara Tonelli and Daniele Pighin. 2009. New features for framenet - wordnet mapping. In *Proceedings of the Thirteenth Conference on Computational Natural Language Learning (CoNLL'09)*, Boulder, CO, USA.