Automatic Induction of FrameNet Lexical Units in Italian

Silvia Brambilla[‡]**, Danilo Croce**[†]**, Fabio Tamburini**[‡]**, Roberto Basili**[†]

[†]University of Rome Tor Vergata, [‡]University of Bologna

{silvia.brambilla2,fabio.tamburini}@unibo.it
{croce,basili}@info.uniroma2.it

Abstract

In this paper we investigate the applicability of automatic methods for frame induction to improve the coverage of IFrameNet, a novel lexical resource based on Frame Semantics in Italian. The experimental evaluations show that the adopted methods based on neural word embeddings pave the way for the assisted development of a large scale lexical resource for our language.

1 Introduction

When dealing with large-scale lexical resources, such as FrameNet (Baker et al., 1998), PropBank (Palmer et al., 2005), VerbNet (Schuler, 2005) or VerbAtlas (Di Fabio et al., 2019), the semi-automatic association between predicates and lexical items (also known as Lexical Units or LUs) is crucial to improve the coverage of a resource while limiting the costs of its manual annotation. Several approaches to this semi-supervised task exist, as discussed in QasemiZadeh et al. (2019). In particular, Pennacchiotti et al. (2008) exploited distributional models of lexical meaning (Sahlgren, 2006; Croce and Previtali, 2010) to induce new LUs consistently with the Frame Semantics theory (Baker et al., 1998), representing words meaning and semantic frames through geometrical word spaces. As a result, this approach allows to induce new LUs when applied to the English version of FrameNet. However, this is a quite consolidated resource with many existing LUs connected to each semantic predicate, i.e., each frame. The applicability of this method in scenarios where only one or two LUs are available for each frame is still an open issue. At the same

time, since the work of Pennacchiotti et al. (2008), the application of neural approaches to the acquisition of word embeddings (Mikolov et al., 2013; Baroni et al., 2014; Ling et al., 2015) significantly improved in terms both of representation capability and scalability of geometrical models of lexical semantics.

In this paper we thus investigate the applicability of the method proposed in Pennacchiotti et al. (2008) to boost the coverage of a novel and still limited lexical resource based on Frame Semantics in Italian. This resource has been developed within the IFrameNet (IFN) project (Basili et al., 2017), which aims at creating a large coverage FrameNet-like resource for Italian and to come up with a complete dictionary in which every lexical entry[1] is linked to all the frames it can evoke (i.e., the frames for which it is a LU). At this moment, while the resource counts more than 7,700 lexical items associated to more than 1,048 frames, each lexical item is connected, on average, to only 1.3 frames, and it is problematic if considering the high polysemy of Italian words (Casadei, 2014).

The experimental evaluation shows that neural word embeddings enable the effective application of the distributional approach from Pennacchiotti et al. (2008) to improve the coverage of IFN. Moreover, the adopted distributional framework allowed to develop a graphical semantic browser to support annotators while assigning new LUs to frames. This study paves the way to the semi-automatic development of IFN and investigates about the applicability of neural word embeddings to the incremental semi-automatic LU induction process.

2 Related Work

In the development of FrameNet and FrameNet-like resources for new languages, one important

---

[1]Where with the term *lexical entry* we denote a lemma, with its Part of Speech tag, that activates at least one LU.

task is the creation of a large-scale dictionary, in order to guarantee an effective application in semantic analyses or NLP tasks. In fact, the limited coverage of FrameNet has been addressed as one of the main reason of failures (Pennacchiotti et al., 2008; Pavlick et al., 2015). For these reasons and given the high costs of manual annotation, both in terms of time and resources (i.e., human annotators), the automatic (or semi-automatic) expansion of the dictionary for FrameNet and FrameNet-like resources has received attention during the years. Several methods to support the population of frames in FrameNet (Baker et al., 2007; Pavlick et al., 2015; Ustalov et al., 2018; QasemiZadeh et al., 2019; Anwar et al., 2019; Arefyev et al., 2019; Yong and Torrent, 2020), and FrameNet-like resources (Johansson and Nugues, 2007; Tonelli et al., 2009; Tonelli, 2010; Johansson, 2014; Hayoun and Elhadad, 2016) with new Lexical Units have been widely investigated. Some of the methodologies proposed in order to automatically expand FrameNet have exploited the alignment between WordNet and FrameNet data (Johansson and Nugues, 2007; Pennacchiotti et al., 2008; Ferrández et al., 2010). Another strategy is the one adopted by Pavlick et al. (2015) where the scholars enlarge FrameNet coverage using automatic paraphrase. The majority of the works dealing with automatic frame induction, however, exploits distributional methods, for example the work on which this research relies the most, i.e., the work of Pennacchiotti et al. (2008) or some of the most recent works such as the ones of Ustalov et al. (2018), Arefyev et al. (2019) and Yong and Torrent (2020). Ustalov et al. (2018), for example, model the frame induction problem as a tri-clustering problem and use dependency triples automatically extracted from a Web-scale corpus. Arefyev et al. (2019) propose to combine dense representations from hidden layers of a masked language model with sparse representations based on substitutes for the target word in the context for the creation of vector representations.

3   IFrameNet status

The IFrameNet project (Basili et al., 2017), relied, as a starting point, on the achievements of previous researches on the development of Italian resources annotated according to Frame Semantics (Tonelli and Pianta, 2009; DeCao et al., 2010), i.e., a set of automatically induced LUs that were covering 554 frames of the $1,224$ frames in FrameNet.

Since the beginning, our main objective has been to improve the coverage of the resource in terms of annotated frames, increasing the number of the LUs and the number of annotated sentences representing each predicate. Starting from the results achieved in 2017, we enlarged the dictionary and provided an initial set of LUs for those frames without any annotation. We also revised the whole dictionary and expunged the LUs whose lemma had low frequency[2] in CORIS (Corpus di Italiano Scritto) (Rossini Favretti et al., 2002). Since CORIS is a large-scale and general-purpose Italian corpus (without biases to any domain), we speculate that not represented LUs can hardly characterize a frame in Italian. Moreover, we worked on the frame annotation of sample sentences taken from the CORIS corpus. We relied on CORIS because it is domain independent and suitable to represent the generic notion of frames. Currently, the resource contains:

- **7,776 lexical entries** of which: $1,130$ adjectives, $4,309$ nouns and $2,337$ verbs;

- **10,379 LUs** (nouns, verbs and adjectives) validated in terms of pairs of lexical entries and evoked frame(s);

- **1,048 frames with at least one LU** among which **743 frames are represented with at least one sentence**. Among the 176 frames that still do not have any LU in their dictionary, 134 are marked as Non-Lexical in FrameNet, 12 do not have any LU in FrameNet, but are not explicitly marked as Non-Lexical, 18 are not represented in FrameNet by any noun, verb or adjective and finally, for just 8 frames, it was difficult to find LUs in Italian (e.g. IMPROVISED_EXPLOSIVE_DEVICE or SHORT_SELLING);

- **5,208 sentences** annotated and validated with at least one LU;

- an average of **9.9 LUs assigned to each frame**;

- an average of **1.3 frames associated to each LU**. Among the existing LUs, $5,960$ are assigned to only one frame. Given that Italian language is highly polysemous, it is probable that many LUs evoke more than one frame. This work aims at reducing this limitation.

---

[2]Less than 20 occurrences in the corpus.

4 Automatic Frame Induction

For the Frame Induction we rely on distributional methods as in Pennacchiotti et al. (2008), described hereafter.

Distributional representation. As a first step, we obtain a distributional representation of the CORIS corpus and represent in the wordspace each LU as a vector $\vec{l}$. We investigated three slightly different approaches for the acquisition of the wordspaces: the Continuous Bag-of-Words model (`CBOW`), the `Skip-gram` model (Mikolov et al., 2013) and the Structured Skip-gram (`sskip-gram`) model (Ling et al., 2015). The `sskip-gram` is a modification of the `skip-gram` model, sensitive to the positioning of the words and, thus, more suitable for capturing syntactic properties of the words (Ling et al., 2015). Our hypothesis is that this last model would be more suitable for capturing LUs frame properties since syntax is, in general, in agreement with semantic arguments (i.e., Frame Elements, FEs) and their order.

"Framehood" representation. As a second step, we exploit the obtained embeddings to represent the meaning of frames. We assume that a frame f can be described by the set of its LUs $l \in F$ and that LUs vectors $\vec{l}$ can be thus used to acquire a distributional representation for each frame. In a nutshell, for each frame we: (i) select all the LUs of its dictionary, (ii) apply to LUs vectors $\vec{l}$ a clustering algorithm. A frame will be then represented as a set of clusters: given that each frame can have various nuances and that it can be representative of non overlapping senses, sparse in the semantic space, we represent it through its "clusters of senses". This captures, in the semantic space, the possible "framehood" distributions, as dense regions of LUs. In this work, we applied standard K-means (Hartigan and Wong, 1979), so that each frame is represented as a set of k clusters. For each frame k is empirically set to the square root of the number of LUs l in that frame: $k = \sqrt{|l|}$, where $|l|$ denotes the count of l per frame. In this way, each f will have k clusters depending on the number of its LUs and the centroid of each cluster will represent the prototype for a subset of the senses of a frame.

New LU induction. Once obtained the distributional representations for frames and LUs, the third step involves the automatic induction of frames given a candidate lexical item. For each

POS	1	2	5
a	295	207	65
n	631	463	250
v	675	514	245
a-n-v	1,041	916	511

Table 1: Number of frames considered according to different filtering policies. In column the threshold applied to the number of required LUs.

candidate predicate word, we computed the distance between its vector and the sets of clusters representing the frames. The "nearest" clusters will be the ones containing a set of LUs more closely related to the input lexical item, so that the corresponding frames will be suggested as its evoking frames.

5 Experimental Evaluation

In order to assess the quality of the proposed method, we evaluate its capability in rediscovering the frames manually associated to a lexical item. We apply a leave-one-out schema: for each candidate lexical item, we eliminate it from the dictionary and query the model to "suggest" up to 10 frames. In practice, we rebuild the clusters and then compute the distance between the lexical item's vector and the set of clusters representing all frames. Then, we compare the suggested frames with the frames that were originally linked to the LU. As in Pennacchiotti et al. (2008), we compute Accuracy as the fraction of LUs that are correctly re-assigned to the original frame. Accuracy is computed at different levels b: a LU is correctly assigned if one of its gold standard frames appears among the best-b frames ranked by the model. In fact, as LUs can have more than one correct frame, we deem as "correct" an assignment for which at least one of the correct frames is among the best-b.

The model is evaluated by sampling the test bed according two dimensions, as reported in Table 1. First, we considered the Part-of-Speech (POS) of the LUs (i.e., rows in Table 1). In fact, lexical items having different POS are generally projected in different sub-spaces within word spaces. We thus evaluate the model considering separately LUs and frames containing adjectives (a), nouns (n) or verbs (v). For the sake of completeness, we also evaluated the model without any selection by POS (row *a-n-v*). When a frame does not contain any LU represented in the wordspace with a required POS, it is discarded during the evaluation: as an example, the actual dictionary contains 631

POS	th	$b-1$	$b-2$	$b-3$	$b-4$	$b-5$	$b-6$	$b-7$	$b-8$	$b-9$	$b-10$
a	1	32%	41%	47%	50%	52%	53%	55%	56%	57%	58%
	2	41%	54%	62%	65%	68%	70%	72%	74%	75%	76%
	5	60%	75%	83%	87%	89%	90%	91%	93%	94%	95%
n	1	42%	53%	59%	62%	65%	66%	68%	69%	70%	71%
	2	48%	61%	67%	71%	73%	75%	76%	78%	79%	80%
	5	59%	72%	79%	82%	85%	87%	88%	89%	90%	91%
v	1	25%	35%	41%	44%	47%	49%	50%	52%	53%	54%
	2	32%	43%	50%	53%	57%	59%	61%	62%	64%	65%
	5	42%	55%	63%	69%	72%	74%	76%	78%	79%	81%
a-n-v	1	36%	47%	53%	56%	59%	61%	62%	64%	65%	66%
	2	41%	52%	58%	63%	65%	67%	68%	70%	71%	72%
	5	49%	61%	68%	73%	75%	78%	79%	81%	82%	83%

Table 2: Accuracy on LU induction according to the best-b ranking and split according to number and POS tag of LUs.

frames containing at least one noun.

Then, we filtered frames by applying a threshold to the number of LUs a frame should be connected to, in order to be considered (columns in Table 1), as it follows: first, we considered all frames containing at least one LU whose lemma occurred at least 20 times in CORIS, without applying any other restriction (column 1); then we filtered frames with at least 2 valid LUs[3] (column 2); finally we filtered frames with at least 5 valid LUs (column 5). Both filter policies can be combined and the stricter these policies are, the lower the number of frames considered in the evaluation. As a consequence, the Accuracy baseline of a model which randomly assigns LUs to frames depends on the number of selected frames: when no filter is applied (row $a-n-v$ and column 1) a random assignment would achieve $0.09\% = \frac{1}{1,041}$ of Accuracy, or $0.4\% = \frac{1}{250}$ when only frames containing at least 5 nouns are selected.

Table 2 reports the experimental results of a model derived using a `sskip-gram` model (Ling et al., 2015)[4]. If we consider the performance over only nouns (n) we see that, when a reasonable threshold is set (row $th = 2$), in 48% of cases in first position we find one of the original frames evoked by the noun under analysis (column $b-1$). If we consider the first two frames proposed by the system ($b-2$) the Accuracy rises up to 61% and it keeps increasing as we consider more frames. It is impressive if considering that the corresponding random baseline is $0.2\% = \frac{1}{463}$ and $0.4\% = \frac{2}{463}$. If we jointly consider nouns, verbs and adjectives

(a-n-v) the performance is slightly lower: for example, with the same threshold $th = 2$ and considering only two suggested frames ($b-2$) the Accuracy is 61%. It means that, on average, the model capability of assigning LUs (ignoring their POS) to frames is slightly lower. This is confirmed by the general drop obtained when only verbs or adjectives are considered: for verbs, considering only the best suggestion ($b-1$) we measured 25%, if we don't apply any threshold, to 32%, if we consider $th = 2$, to 42% if we consider $th = 5$. This is mainly due to higher polysemy characterizing verbs and adjectives with respect to nouns (Casadei, 2014). Anyway, this result is straightforward if considering that for verbs, the baseline in the setting $th = 2$ and $b = 1$ corresponds to $0.2\% = \frac{1}{514}$.

Discussion. It is worth noting that our dictionary is largely incomplete and thus some of those counted as "incorrect assignments" are instead frames that are evoked by the LU under analysis and that *should be added* to the dictionary. Moreover, we can see that many of the $b-10$ frames are often related at different degrees with the lexical entry under analysis and with the frames for which it is a LU.

For example, when considering the lexical entry "*impiccare.v*" (*hang.v*) the model does not retrieve among the $b-10$ suggestions the only "correct" frame, i.e., the frame EXECUTION. Anyway, the closest frame identified is the frame KILLING that not only is linked with EXECUTION with an Inheritance relation, but also appears to be evoked by "*impiccare.v*". Again, the system is not able to re-assign the lexical entries "*innalzarsi.v*" (*raise.v and rise.v*), "*innocenza.n*" (*innocence.n*) and "*radiazione.n*" (*radiation.n or expulsion.n*) . Anyway, in the $b-10$

---

[3]This threshold also overcomes the intrinsic limitation of the leave-one-out schema; when considering frames with only one LU, it becomes impossible to spot the original frame in the test data because it will not be represented by any LU.

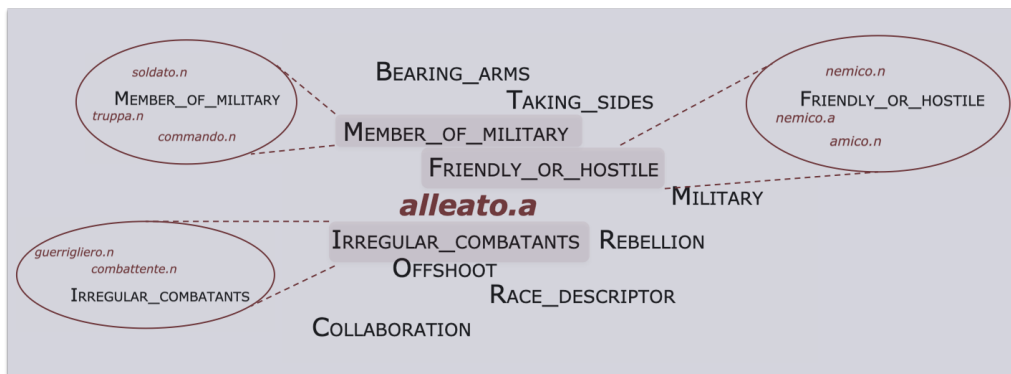[4]This method outperformed the `CBOW` and `skip-gram`, not reported here for lack of space.

Figure 1: An example of the IFrameNet Navigator for the LU *alleato.a*

of "*innalzarsi.v*" appears in fourth position the frame CHANGE_POSITION_ON_A_SCALE that can be evoked by "*innalzarsi.v*" in sentences such as "*La marea si innalzava*" (*The tide was rising*) and in the $b - 10$ of "*innocenza.n*" appears, in first position, the frame CANDIDNESS that is evoked by this LU in sentences such as "*Lei rispose con innocenza*" (*She answered genuinely*). The term "*radiazione.n*" is present in the dictionary only with the meaning *expulsion.n* and it is linked only to EXCLUDE_MEMBER. Nevertheless, the system proposes the frame NUCLEAR_PROCESS in first position and retrieves one correct meaning of a LU like "*radiation.n*". For "*alleato.a*" (*ally.n*, also shown in Figure 1) the system proposes a "correct" frame in ninth position. Anyway, we find in second position the frame MEMBER_OF_MILITARY that can be plausibly evoked. Moreover the LU "*agnello.n*" (*lamb.n*) evokes in the dictionary only the frame FOOD; anyway, as correctly suggested by the system, it is also LU of the frame ANIMALS. Moreover for "*agnello.n*" the system proposes also, in sixth position, PEOPLE_BY_MORALITY that recalls the idea of innocence and righteousness that represents (at least for the Italian language) a metaphorical extension of the meaning of "*lamb.n*", strongly influenced by the religious image of the lamb.

In some other cases, the system suggests relations between frames. For example, if we consider the lexical entry "*identico.a*" (*identical.a* from IDENTICALITY) we see in the best-10 frames that the system proposes frames such as SIMILARITY (first position) or DIVERSITY (seventh position). If we look at the frame-to-frame relations in FrameNet, we see that IDENTICALITY and SIMILARITY or IDENTICALITY and DIVERSITY are not directly connected even if they appear, at a close analysis, strictly related.

6 IFrameNet Navigator

In order to make the model valuable for the annotators, we also developed a Graphical User Interface, called IFrameNet Navigator. It allows querying and navigating the geometrical representation of semantic phenomena as it displays, for each lexical entry in the dictionary, the best-10 frames. These can be also selected to browse the set of LUs assigned to the cluster underlying the frame, as shown in Figure 1. Finally, each LU can be selected to browse the list of corresponding annotated sentences.

The objectives of the Navigator are: (*i*) to support the analysis of the currently modeled lexical entries (and the corresponding LUs); (*ii*) to support the validation of the current sentence classification; (*iii*) the mining of the CORIS corpus for improving the semantic coverage of the resource for the Italian language; (*iv*) in perspective, to offer support towards crowd sourcing.

This tool will be publicly released to trigger collaborative validation and annotation as an extension of the IFrameNet and the CORIS resources.

7 Conclusions and Research Perspectives

In this work, we presented the actual state of the IFrameNet project, which aims at developing a large-scale lexical resource based on Frame Semantics in Italian. Moreover, we investigated the applicability of a method for the automatic Induction of FrameNet Lexical Units to improve the coverage of the actual resource, in terms of number of frames assigned to the almost 8,000 existing lexical entries.

With respect to previous work, i.e., Pennacchiotti et al. (2008) we empirically demonstrate the beneficial impact of neural word embeddings in the overall workflow in Italian. The robustness

of the adopted model is confirmed also when applied to a resource with a limited average number of frames associated to Lexical Units. The experimental evaluations in many cases showed the valuable support of the method in discovering new Lexical Units by suggesting novel evoked frames. Moreover, the error analysis suggested that most of the "discarded" frames still entertain various kinds of relationships with the "correct" ones as defined in FrameNet, such as Inheritance or Usage. In some cases, it also highlighted metaphorical meanings that the lexical entries could assume.

As a future work, we will certainly exploit the produced IFrameNet Navigator to extend the current LU Italian dictionary, support the annotation of novel sentences and introduce frame-to-frame relations in Italian. Another path that might worth investigating is the exploitation of dependency-based word embeddings for the distributional representation of LUs and frames. This may beneficial since dependency-based contexts highlight more functional similarities (Levy and Goldberg, 2014). Finally, we plan to use the derived frame distributions to augment existing contextualized embeddings in support of Frame Induction (Sikos and Padó, 2019) or Semantic Role Labeling (Shi and Lin, 2019) tasks.

References

Saba Anwar, Dmitry Ustalov, Nikolay Arefyev, Simone Paolo Ponzetto, Chris Biemann, and Alexander Panchenko. 2019. Hhmm at semeval-2019 task 2: unsupervised frame induction using contextualized word embeddings. *arXiv preprint arXiv:1905.01739*.

Nikolay Arefyev, Boris Sheludko, Adis Davletov, Dmitry Kharchev, Alex Nevidomsky, and Alexander Panchenko. 2019. Neural granny at semeval-2019 task 2: A combined approach for better modeling of semantic relationships in semantic frame induction. In *Proceedings of the 13th International Workshop on Semantic Evaluation*, pages 31–38.

Collin F. Baker, Charles J. Fillmore, and John B. Lowe. 1998. The Berkeley FrameNet project. In *Proc. of COLING-ACL*, Montreal, Canada.

Collin F Baker, Michael Ellsworth, and Katrin Erk. 2007. Semeval-2007 task 19: Frame semantic structure extraction. In *Proceedings of the Fourth International Workshop on Semantic Evaluations (SemEval-2007)*, pages 99–104.

Marco Baroni, Georgiana Dinu, and Germán Kruszewski. 2014. Don't count, predict! a systematic comparison of context-counting vs. context-predicting semantic vectors. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 238–247, Baltimore, Maryland, June. Association for Computational Linguistics.

Roberto Basili, Silvia Brambilla, Danilo Croce, and Fabio Tamburini. 2017. Developing a large scale framenet for italian: the iframenet experience. *CLiC-it 2017 11-12 December 2017, Rome*, page 59.

Federica Casadei. 2014. La polisemia nel vocabolario di base dell'italiano. *Lingue e Linguaggi*, 12:35–52.

Danilo Croce and Daniele Previtali. 2010. Manifold learning for the semi-supervised induction of FrameNet predicates: An empirical investigation. In *Proceedings of the 2010 Workshop on GEometrical Models of Natural Language Semantics*, pages 7–16, Uppsala, Sweden, July. Association for Computational Linguistics.

Diego DeCao, Danilo Croce, and Roberto Basili. 2010. Extensive evaluation of a framenet-wordnet mapping resource. In Nicoletta Calzolari (Conference Chair), Khalid Choukri, Bente Maegaard, Joseph Mariani, Jan Odijk, Stelios Piperidis, Mike Rosner, and Daniel Tapias, editors, *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*, Valletta, Malta, may. European Language Resources Association (ELRA).

Andrea Di Fabio, Simone Conia, and Roberto Navigli. 2019. Verbatlas: a novel large-scale verbal semantic resource and its application to semantic role labeling. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 627–637.

Oscar Ferrández, Michael Ellsworth, Rafael Munoz, and Collin F Baker. 2010. Aligning framenet and wordnet based on semantic neighborhoods. In *LREC*, volume 10, pages 310–314.

J. A. Hartigan and M. A. Wong. 1979. A k-means clustering algorithm. *JSTOR: Applied Statistics*, 28(1):100–108.

Avi Hayoun and Michael Elhadad. 2016. The hebrew framenet project. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 4341–4347.

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*, pages 27–30.

Richard Johansson. 2014. Automatic expansion of the swedish framenet lexicon: Comparing and combining lexicon-based and corpus-based methods. *Constructions and Frames*, 6(1):92–113.

Omer Levy and Yoav Goldberg. 2014. Dependency-based word embeddings. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 302–308.

Wang Ling, Chris Dyer, Alan W Black, and Isabel Trancoso. 2015. Two/too simple adaptations of word2vec for syntax problems. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1299–1304.

Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. Distributed Representations of Words and Phrases and their Compositionality. In C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 26*, pages 3111–3119. Curran Associates, Inc.

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

Ellie Pavlick, Travis Wolfe, Pushpendre Rastogi, Chris Callison-Burch, Mark Dredze, and Benjamin Van Durme. 2015. Framenet+: Fast paraphrastic tripling of framenet. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pages 408–413.

Marco Pennacchiotti, Diego De Cao, Roberto Basili, Danilo Croce, and Michael Roth. 2008. Automatic induction of framenet lexical units. In *Proceedings of the 2008 conference on empirical methods in natural language processing*, pages 457–465.

Behrang QasemiZadeh, Miriam R. L. Petruck, Regina Stodden, Laura Kallmeyer, and Marie Candito. 2019. SemEval-2019 task 2: Unsupervised lexical frame induction. In *Proceedings of the 13th International Workshop on Semantic Evaluation*, pages 16–30, Minneapolis, Minnesota, USA, June. Association for Computational Linguistics.

Rema Rossini Favretti, Fabio Tamburini, and Cristiana De Santis. 2002. Coris/codis: A corpus of written italian based on a defined and a dynamic model. *A rainbow of corpora: Corpus linguistics and the languages of the world*, pages 27–38.

Magnus Sahlgren. 2006. *The Word-Space Model*. Ph.D. thesis, Stockholm University.

Karin Kipper Schuler. 2005. *VerbNet: A broad-coverage, comprehensive verb lexicon*. Ph.D. thesis, University of Pennsylyania.

Peng Shi and Jimmy Lin. 2019. Simple BERT models for relation extraction and semantic role labeling. *CoRR*, abs/1904.05255.

Jennifer Sikos and Sebastian Padó. 2019. Frame identification as categorization: Exemplars vs prototypes in embeddingland. In *Proceedings of the 13th International Conference on Computational Semantics - Long Papers*, pages 295–306, Gothenburg, Sweden, May. Association for Computational Linguistics.

Sara Tonelli and Emanuele Pianta. 2009. Three issues in cross-language frame information transfer. In *Proceedings of the International Conference RANLP-2009*, pages 441–448, Borovets, Bulgaria, September. Association for Computational Linguistics.

Sara Tonelli, Daniele Pighin, Claudio Giuliano, and Emanuele Pianta. 2009. Semi-automatic development of framenet for italian. In *Proceedings of the FrameNet Workshop and Masterclass, Milano, Italy*.

Sara Tonelli. 2010. *Semi-automatic techniques for extending the FrameNet lexical database to new languages*. Ph.D. thesis, Università Ca'Foscari Venezia.

Dmitry Ustalov, Alexander Panchenko, Andrei Kutuzov, Chris Biemann, and Simone Paolo Ponzetto. 2018. Unsupervised semantic frame induction using triclustering. *arXiv preprint arXiv:1805.04715*.

Zheng Xin Yong and Tiago Timponi Torrent. 2020. Semi-supervised deep embedded clustering with anomaly detection for semantic frame induction. In *Proceedings of The 12th Language Resources and Evaluation Conference*, pages 3509–3519.