HAL
archives-ouvertes.fr

Unsupervised extraction of semantic relations using discourse cues

Juliette Conrath, Stergos Afantenos, Nicholas Asher, Philippe Muller

▶ To cite this version:

HAL Id: hal-01142604
https://hal.archives-ouvertes.fr/hal-01142604

Submitted on 15 Apr 2015

Unsupervised extraction of semantic relations using discourse cues

Juliette Conrath   Stergos Afantenos   Nicholas Asher   Philippe Muller

IRIT, Université Toulouse & CNRS, Univ. Paul Sabatier, 118 Route de Narbonne, 31062 Toulouse

{firstname.lastname@irit.fr}

Abstract

This paper presents a knowledge base containing triples involving pairs of verbs associated with semantic or discourse relations. The relations in these triples are marked by discourse connectors between two adjacent instances of the verbs in the triple in the large French corpus, frWaC. We detail several measures that evaluate the relevance of the triples and the strength of their association. We use manual annotations to evaluate our method, and also study the coverage of our resource with respect to the discourse annotated corpus Annodis. Our positive results show the potential impact of our resource for discourse analysis tasks as well as other semantically oriented tasks like temporal and causal information extraction.

1 Introduction

Relational lexical resources, which describe semantic relations between lexical items, have traditionally focused on relations like synonymy or similarity in thesauri, perhaps including some hierarchical semantic relations like hyperonymy or hyponomy or part-whole relations as in the resource Wordnet (Felbaum, 1998). Some distributional thesauri contain more varied relations, see e.g. (Grefenstette, 1994), however these relations are not typed. The lexical semantics given by FrameNet (Baker et al., 1998) does include causal and temporal relations, as does Verbocean (Chklovski and Pantel, 2004), but coverage is limited and empirical validation of these resources is partial and still largely remains to be done.

Lexical relations, in particular between verbs, are nevertheless crucial for understanding natural language and for many information processing tasks. They are needed for textual inference, in which one has to infer certain relations between eventualities (Hashimoto et al., 2009; Tremper and Frank, 2013), for information extraction tasks, like finding temporal relations between eventualities mentioned in a text (UzZaman et al., 2013), for automatic summarization (Liu et al., 2007), and for discourse parsing in the absence of explicit discourse markers (Sporleder and Lascarides, 2008).

In this paper we report on our efforts to extract semantic relations essential to the analysis of discourse and its interpretation, in which links are made between units of text or rather their semantic representations as in (1) in virtue of semantic information about the two main verbs of those clauses.

(1) The candidate <u>demonstrated</u> his expertise during the interview. The committee was completely <u>convinced</u>.

We follow similar work on the extraction of causal, temporal, entailment and presuppositional relations from corpora (Do et al., 2011; Chambers and Jurafsky, 2008; Hashimoto et al., 2009; Tremper and Frank, 2013), though our goals and validation methods are different. While one of our goals is to use this information to improve performance in predicting discourse relations between clauses, we believe that such a lexical resource will have other uses in other tasks in which semantic information is needed.

Discourse analysis is a difficult task. Rhetorical relations are frequently implicit and require for their identification inference using diverse sources of lexical and compositional semantic information. In the Penn Discourse Treebank corpus for example, 52% of the discourse relations are unmarked (Prasad et

al., 2008). Accordingly, annotation with discourse structure is a slow and error prone task, and relatively little annotated data is currently available ; and so machine learning approaches have had limited success in this area. Our approach addresses this problem, using non annotated data with features that can be automatically detected to find typical contexts (pairs of discourse units) in which various discourse relations occur. We suppose with (Sporleder and Lascarides, 2008; Braud and Denis, 2013) that such contexts display regular lexical associations, in particular with verbs in those discourse units. An explicit, manually compiled list of all possible associations between two verbs and the semantic relations they suggest is infeasible, so we present here an automatic method for compiling such a list, inspired by the Verbocean project (Chklovski and Pantel, 2004).

Our hypothesis, supported by existing corpora, is that adjacent clauses are often arguments of discourse relations. When these clauses contain certain adverbs or other discourse connectors, we can recover automatically one or more discourse relations that we associate with the main verbs of those clauses. We extract triples consisting of the two verbs and a semantic relation from a large corpus with the aim of inferring that such a pair of verbs can suggest the semantic relation even in the absence of an explicit discourse marker. We thus also suppose, with (Sporleder and Lascarides, 2008; Braud and Denis, 2013), that such discourse markers are at least partially redundant ; inferring a discourse relation between two clauses relies not only the marker but on the two verbs in the related clauses as well. All of our work has been done on French data.

Our paper is organized as follows. We describe first the knowledge base of verb semantic relation triples that we have constructed (section 2) ; we then present our methods for isolating verb pairs implicating discourse or temporal information (section 3). A third section describes our methods of evaluation (section 4) and a fourth discusses related work (section 5).

2   Exploring relations between verbs in a corpus

We built a knowledge base (V^2R) [1] using the frWaC corpus(Baroni et al., 2009). frWaC contains about 1.6 billion words and was collected on the Web on the .fr domain. We first parsed the documents in our corpus using BONSAI [2], which first produced a morpho-syntactic labeling using MElt (Denis and Sagot, 2012) and then a syntactic analysis in the form of dependency trees via a French version of the MaltParser (Nivre et al., 2007).

Our goal is to find pairs of verbs linked by a relation explicitly marked by a discourse connector in the corpus, as an indication of a regular semantic relation between the two verbs. The relations we have considered are common to most theories of discourse analysis, and they can be grouped into four classes (Prasad et al., 2008) : causal (*contingency*) relations, temporal relations, comparison relations (mainly contrast type relations), and expansion relations (e.g. elaboration or continuation).

To find explicitly marked relations, we used a lexicon of discourse connectors for French, the manually constructed LEXCONN resource (Roze et al., 2012) [3]. LEXCONN includes 358 connectors and gives their syntactic category as well as associated discourse relations inspired from (Asher and Lascarides, 2003). Some connectors are ambiguous in that they are associated with several relations. We used only the unambiguous connectors (263 in all) in LEXCONN, as a first step. We regrouped the LEXCONN relations into classes [4] : explanation relations (*parce que*/because) and result (*ainsi*/thus) form the causal class ; temporal relations (*puis, après que*/then,after that) form the narration group. We also considered other relations like contrast (*mais*/but), continuation (*et, encore*/and,again), background (*alors que*/while), temporal location (*quand, pendant que*/when), detachment (*de toutes façons*/anyway), elaboration (*en particulier*/in particular), alternation (*ou*/or), commentary (*au fait*/by the way), rephrasing (*du moins*/at least), and evidence (*effectivement*/indeed).

We searched our syntactically parsed corpus for connectors. When a connector is found and its syntactic category verified, if it is close enough to the root of the sentence (at most one dependency link from the root), we look for an inter-sentential link. The first verb of our pair corresponds in this case

---

1. Available as an SQLite database at `https://dl.dropboxusercontent.com/u/78938139/v2r_db`
2. `http://alpage.inria.fr/statgram/frdep/fr_stat_dep_parsing.html` or (Candito et al., 2010)
3. Freely available at : `https://gforge.inria.fr/frs/download.php/31052/lexconn.tar.gz`.
4. We illustrate each relation with examples of potentially ambiguous markers.

to the last verb of the previous sentence in the case of connectors for narration, or to its main verb for all the other relations. We search for the second verb in the pair within a window of two dependency links after the connector. If the connector is not close enough to the root of the sentence, we look for a intra-sentential link. In this case, we look for the two verbs of the pair in the same sentence within a forward and backward window of two dependency links.

If two verbs are found, we examine their local context to better characterize their usage and to improve our results. If one of the verbs is a modal or support verb, we look for the verb dependent on the modal or support verb and use that as the verb in our pair (if it exists), while keeping the presence of the support verb in memory. Unlike support verbs, we use the presence of a negation or a reflexive particle in the local context to distinguish verbs with different meanings ; e.g., *comprendre*/understand vs. *ne pas comprendre*/not understand, *agir*/act vs. *s'agir*/concern are all distinct entries. To get at different verb senses, we search for idiomatic usage of prepositions using the Dicovalence resource (Van Den Eynde and Mertens, 2010), which contains valency frames for more than 3700 simple French verbs. We also use the Lefff resource (Sagot, 2010) to find idiomatic verbal locutions. We also encode other information that do not lead to distinct lexical entries : tense, and voice.

Relation	Distribution
contrast	50,104%
cause	33,108%
continuation	8,243%
narration	6,362%
background	1,853%
temporal localisation	0.177%
detachement	0.149%
elaboration	0.002%
alternation	0.002%

TABLE 1 – Distribution of relations in V^2R ;commentary, reformulation and evidence occur with negligible frequency.

Once we have obtained a list of verb pairs associated with a connector, we aggregate this data to get a list of triple types (verb1, verb2, relation). Given that we have used only unambiguous connectors (so classified by LEXCONN), the association of a relation with a connector is immediate. We associate to each triple type the number of intra-sentential, inter-sentential and total number of occurrences. The other features mentioned above are stored in a separate table.

Our method has isolated more than 1 million distinct types of triples for V^2R and 2 million occurrences, of which 95% are intra-sentential[5]. Among these triples, 6.2% have 5 or more occurrences.

Table 1 summarizes the distribution of triples by relation in V^2R. Note that triples with contrast and causal relations comprise the majority. This does not mean that these are the most frequent relations in the corpus but only that they are the most frequently marked by the connectors we considered. This makes for a very different distribution than that of the French manually annotated discourse corpus Annodis (Afantenos et al., 2012).

3 Measuring the association of a pair of verbs with a relation

In the last section we presented our extraction method. We now present the measures we have used to rank verb pairs with respect to the strength of their association with a particular discourse relation. We adapted versions of standard lexical association measures like PMI (*pointwise mutual information*) and their variants, as well as some measures specific to the association of a causal relation between items (Do et al., 2011). We also experimented with a new measure specifically designed for our knowledge base.

Measures of lexical association used in research on co-occurrences in distributional semantics pick out significant associations, taking into account the frequency of the related items. We examined over 10 measures ; we discuss the ones with the best results (see section 4). One simple measure, PMI, and its variants, *normalized*, *local* (Evert, 2005), *discounted* (Lin and Pantel, 2002), which are designed to reduce biases in the original measure, work well. The idea behind PMI is to estimate whether the probability of the co-occurrence of two items is greater than the *a priori* probability of the two items appearing independently. In distributional semantics, the measure is also used to estimate the significance of two items co-occurring with a particular grammatical dependency relation like the subject or object relation between an NP and a verb. This use of PMI measures over triples in distributional semantics fits perfectly with our task of measuring the significance of triples consisting of a pair of verbs and

---

5. The low proportion of inter-sentential occurrences comes from our conservative scheme for finding these occurrences, which uses only those connectors at the beginning of the second sentence. Other schemes are possible but would, we fear, introduce too much noise into the data.

a particular semantic or discourse relation ; our PMI measures estimate whether the co-occurrence of two items with a particular discourse relation is higher than the *a priori* probability of the three items occurring independently. Our measures consider co-occurrences of two lexical items in a certain relation denoted by an explicit discourse marker. PMI and normalized PMI are defined as :

$$PMI = \log(\frac{P(V_1, V_2, R)}{P(V_1) \times P(V_2) \times P(R)}) \qquad PMI\_normalized = \frac{PMI}{-2\log(P(V_1, V_2, R))}$$

Indeed, when we have a complete co-occurrence of the three items, we have : $P(V_1) = P(V_2) = P(R) = P(V_1, V_2, R)$, and $PMI = -2\log(P(V_1, V_2, R))$. The values of normalized PMI lie between -1 and 1, approaching -1 when the items never appear together, taking the value 0 in the case of independence, and the value 1 when they always appear together. We also considered a weighted PMI measure (Lin and Pantel, 2002) that corrects the bias of PMI for rare triples.

A specificity measure (Mirroshandel et al., 2013), originally used to measure the precision of subcategorization frames, also performed well :

$$specificity = \frac{1}{3} \times (\frac{P(V_1, V_2, R)}{\sum_i P(V_1, V_i, R)} + \frac{P(V_1, V_2, R)}{\sum_i P(V_i, V_2, R)} + \frac{P(V_1, V_2, R)}{\sum_i P(V_1, V_2, R_i)})$$

A version of Do et al. (2011)'s measure for triples involving causal relations did not fare so well on other types of relation. The definition of the measure can be found in (Do et al., 2011). [6]

Finally, we investigated a measure that evaluates the contribution of each element in the triple to the significance measure (this measure is similar to specificity).

$$W_{combined}(V_1, V_2, R) = \frac{1}{3}(w_{V_1} + w_{V_2} + w_R)$$

with : $w_{V_1} = \frac{P(V_1, V_2, R)}{\max_i(P(V_i, V_2, R))}, w_{V_2} = \frac{P(V_1, V_2, R)}{\max_i(P(V_1, V_i, R))}$, and $w_R = \frac{P(V_1, V_2, R)}{\max_i(P(V_1, V_2, R_i))}$.

4 Evaluating extracted relations

We evaluated V^2R in several ways ; we provided : (i) an intrinsic evaluation of the relations between verbs (section 4.1) and (ii) an extrinsic evaluation where we evaluated the coverage of the resource on a discourse annotated corpus and its potential to help in predicting discourse relations in contexts with no explicit marking (section 4.2).

4.1 Intrinsic evaluation

Our intrinsic evaluation first evaluates the feasibility of assigning an "inherent" semantic link to a verb pair, independently of any linguistic context. For example, is it possible to judge that there is a typical causality link between *push* and *fall*, in scenarios where they share some arguments (subject, object, ...), these scenarios being left to the annotator's imagination (section 4.1.1). In a second stage, we selected several verb pairs linked with different relations in V^2R, and 40 contexts in which these verbs occur together in the original corpus, to judge the semantic link in context (section 4.1.2).

In both cases we restricted the study to three relation groups : causal, contrastive, and narrative. These are the most often marked relations and correspond to different types of links with a meaningful semantic aspect (as opposed to the "continuation" relation for instance, which is often marked too).

4.1.1 Out of context evaluation

For out of context judgments, we adopted the following protocol : one of the authors chose for each relation 100 verbs with equivalent proportions of good and bad normalized PMI scores. Then the other

---

6. We simplified their measure by ignoring IDF (inverse document frequency) and the distance between the verbs, as neither measure applies to our task.

three authors judged the validity of associating each of the 300 pairs with the corresponding relation, without any knowledge of the source of these pairs.

We measured the inter-annotator agreements with Cohen's Kappa (Carletta, 1996), which resulted in : 0.17 for cause, 0.42 for narration and 0.56 for contrast as mean values. If a 0.6 kappa serves a measure for a feasible semantic judgment task, out of context judgments appear very difficult, with only contrastive pairs as a relative exception. We decided to only consider judgments about contrast, after an adjudication phase, and we evaluated the measures presented in section 3 to see if they could discriminate between the two verb groups, those judged positively or negatively according to human annotations. A Mann-Whitney U statistical test showed all of our measures to be discriminative, with the exception of raw co-occurrence counts for which $p > 0.05$.

4.1.2 In context evaluation

We also judged associations in context. This task was easier and also gave more fine-grained results, because with it we can quantify the degree of association, and the typicality of the link, as a proportion of contexts where the two verbs appear together in a given semantic relation. We can then observe if this proportion is correlated with the association measures we already presented. Nevertheless, this is a costly way of evaluating a verb pair, as we require a number of judgments on each pair. It is also not easy to sample the possible pairs with different values to be able to observe significant correlations, because we cannot predict in advance how they will be judged by the annotators.

We selected 40 contexts for each of the 15 pairs of verbs we chose, 5 for each of the target relation (cause, narration, contrast). Selected pairs range over different values of normalized PMI, again chosen by one of the authors independently of the others, who annotated the 600 contexts. Prior to adjudication, raw agreement was 78% on average, for an average kappa of 0.46 (and a maximum of 0.49). These values seem moderately good, as the task is also rather difficult.

Verb pair	translation	association /human
Cause		
inviter/souhaiter	invite/wish	12.8%
promettre/élire	promise/elect	25.6%
aimer/trouver	like/find	38.5%
bénéficier/créer	benefit/create	51.3%
aider/gagner	help/win	53.8%
Contrast		
proposer/refuser	propose/refuse	59.0%
augmenter/diminuer	increase/decrease	64.1%
tenter/échouer	try/fail	64.1%
gagner/perdre	win/lose	71.8%
autoriser/interdire	authorize/forbid	74.4%
Narration		
parler/réfléchir	speak/think	42.5%
acheter/essayer	buy/try	70.0%
atteindre/traverser	reach/cross	77.5%
commencer/finir	begin/end	80.0%
envoyer/transmettre	send/transmit	82.5%

TABLE 2 – For each relation, the list of verb pairs manually evaluated in context (and an approximate translation), and the association percentage resulting from the adjudicated human annotation.

Table 2 shows the results after adjudication : for each pair, the proportion of contexts in which the considered relation is judged to appear.

We computed two correlation values between the association ratio in contexts manually annotated and each association measure considered : one based on all annotated contexts, and one on the subset of contexts devoid of explicit markers of a semantic relation (implicit contexts). The latter is important to quantify the actual impact of the method, since explicit marking is already used as the basis of verb association in the same corpus. Implicit contexts, however, never appeared in the computation of the verb pair associations.

	normalized PMI	specificity	W_combined	discounted PMI	PMI	local PMI	U_do	raw frequency
Global correlation	0.749	0.747	0.720	0.716	0.709	0.434	0.376	0.170
Correlation for implicit instances	0.806	0.760	0.738	0.761	0.756	0.553	0.499	0.242

TABLE 3 – Pearson correlation for the 15 pairs considered and measures from section 3, in decreasing order.

Table 3 shows that mutual information measures are well correlated with human annotations, and that our W_combined seems useful too. We also observed results on each relation separately, although one should be careful drawing conclusions from these results since the correlations are then computed on 5 points only. These results (not shown here) show a lot of variation between relations. The U_do measure, designed for causal relations, does indeed produce good results for these relations, but does not generalize well to our other chosen relations.

Also, local PMI seems to work very well on narration and causal relations. This needs to be confirmed with more verb pairs.

We conclude that the best three measures are : normalized PMI, specificity, and W_combined. The last two assign their maximal value to several pairs, so we used them in a lexicographical ordering to sort all associated pairs, using normalized PMI to break ties.

Verb pair	Translation	Relation
abandonner / mener	abandon / lead	background
ne pas s'arrêter / rouler	not stop / drive	narration
donner satisfaction sur / réélire	give satisfaction concerning / re-elect	continuation
emporter / ne pas cesser	take away / not stop	summary
emprunter / assurer	borrow / insure	cause
ne pas manquer / prolonger	not miss / prolong	detachment
ratifier / trembler	ratify / tremble	background
avoir honte / faire pitié	be ashamed / cause pity	cause
avoir droit / cotiser pour	be entitled / contribute to	temploc
ne pas représenter / stéréotyper	not represent / stereotype	temploc

TABLE 4 – Ten best triples in the database.

Table 4 shows the best triples with our lexicographical ranking.

4.2 Extrinsic evaluation

In order to evaluate the performance of our resource relative to its main intended application—predicting rhetorical relations in text, we intend to use our association measures as additional features to an inductive prediction model. Whether this evaluation produces results depends on the proportion of cases in which this information could help and on the coverage of our resource with respect to these cases. We used the Annodis corpus (Afantenos et al., 2012), a set of French texts annotated with rhetorical relations, for our study.

To improve existing models, a significant number of the predictions to be made must involve a verb pair for which we have information in the resource. A first indication of its usefulness is also that the verb pair appears most frequently with the relation group to which the annotation belongs, for instance the fact that two verbs are related with a causal relation whenever we want to predict an explanation. This is interesting only in the absence of an explicit marking of the target relation, i.e for implicit relations.

Beyond that, it should be interesting to use all the available information about other semantic relations too : for instance a potential causal link between two events could indicate the relevance of a temporal link for the prediction of a relation. We relied again on the Lexconn marker database. As an approximation we considered that a relation between two discourse units is explicit when a Lexconn marker is present in any of the two segments, and one of the potential senses of the marker is the annotated relation. This may overestimate the number of explicit instances but ensures that all implicit instances are indeed implicit (assuming a good enough coverage of the marker resource). The Annodis corpus lists rhetorical relations between elementary discourse units (EDUs), typically clauses, and complex discourse units (sets of EDUs) ; as a simplification we only consider EDUs, since the question of what is a main verb of a complex unit is difficult to answer. This is a relatively small corpus, as it includes about 2000 instances of relations between elementary discourse units.

Table 5 present results for coverage, for the main relations in the annotated corpus. Note that only a small part of the set of relations between EDUs is considered when we restrict instances to both EDUs with verbs (about 20% of the whole). It turns out that a lot of EDUs in Annodis are short segments (incises, detached segments, ...).

	global	narration	cause	contrast	elab.	cont.	BG	other
Annodis pairs	427	73	67	41	96	92	24	16
Annodis pairs $\in V^2R$	68.9	71.2	70.8	78.0	68.3	61.9	74.1	62.5
Annodis triples $\in V^2R$	26.5	34.2	50.0	70.7	0.0	20.6	11.1	0.0
Implicit Annodis pairs	83.4	71.2	79.2	36.6	99.0	94.8	88.9	100.0
Implicit Annodis pairs $\in V^2R$ (any relation)	56.9	52.1	54.2	31.7	67.3	58.8	66.7	62.5
Implicit Annodis triples $\in V^2R$ (with correct relation)	17.7	24.7	40.3	31.7	0.0	19.6	11.1	0.0

TABLE 5 – Coverage of verb pairs in V^2R with respect to EDU pairs in the Annodis corpus containing two verbs. Except for the first line, all numbers are percentages. Pair = verb pairs in the EDUs linked by a rhetorical relation R, Triple=verb pair associated with a relation R in V^2R, BG = Background, cont.=continuation, elab.=elaboration.

Our table includes : the proportion of verb pairs found in Annodis EDUs that appear in V^2R, the proportion of triples from Annodis that appear in V^2R (with the correct relation), and the restriction of these proportions to implicit contexts in Annodis. Except for a few exceptions due to lemmatisation errors, all verbs in Annodis are in V^2R in at least one pair, and we can see that the pairs in V^2R cover most of the pairs appearing in Annodis (almost 70% globally and between 60 and 80% depending on the relation), and a little less of implicit cases (around 55% on average). We note that a high proportion of the implicit cases contains verb pairs that have been collected in a marked context, even for rarely marked relations like elaboration or continuation—contexts with these relations are the majority in Annodis. Furthermore more than half of these contexts are associated with the right relation in V^2R. Thus the hypothesis of the partial redundancy of connectors appears useful when isolating verbal associations relevant for discourse from a large corpus. We also looked at semantic neighbors of the verbs in V^2R but this did not increase coverage significantly.

A good test of the predictive power of the semantic information we gathered is also to include the association measures as additional features to a predictive model, to improve classically low results on implicit discourse relations. The only available discursive corpus in French, Annodis, is small, and as shown above only about 400 instances have a verb in both related EDUs. We trained and tested a maximum entropy model with and without the association measures as features, on top of features presented in Muller et al. (2012), who trained a relation model on the same corpus. We did a 10-fold cross-validation on the 400 instance subset as evaluation, and did not find a significant difference between the two set-ups (F1 score was in the range .40–.42, similar to the cited paper), which is unsurprising

given the size of the subset. We plan to evaluate our method relative to discourse parsing by building an English resource like V²R ; we will then be able to use the much larger PDTB corpus (10 times as large as Annodis) as a source of implicit discourse relations. This should prove a much more telling evaluation of the usefulness of association measures in predicting implicit discourse relations.

5   Related work

There are two different groups of related work. The first group aims to alleviate the lack of annotated data for discourse parsing by using a weakly supervised approach, exploiting the presence of discourse connectors in a large non-annotated corpus. Each pair of elementary discourse units is automatically annotated with the discourse relation triggered by the presence of the connector (connectors are often filtered for non-discursive uses). Those connectors are afterwards eliminated from the corpus so that the model trained on this dataset will not be informed by the presence of those connectors. The pioneering article in this group is Marcu and Echihabi (2002). Such learning methods with such "artificial data" obtain low scores, barely above chance as shown in Sporleder and Lascarides (2008). Braud and Denis (2013) observe that the performance of a classifier for the prediction of implicit relations is much lower when using "artificial" data than on "natural" data (implicit relations annotated by a human being). They propose a method which exploits these two different kinds of datasets together in various mixtures and on the level of the prediction algorithm, obtaining thus a significant improvement on the Annodis corpus. Our approach is different and complementary ; we isolate the semantic relations between pairs of verbs. We can use that as a feature on discourse units for discourse parsing but it has other uses as well.

A second group aims at identifying discourse relations (implicit or not) by focusing on the use of fine-grained lexical relations as another feature during the training phase. Most of this work focuses mainly on the use of lexical relations between two verbs. Chklovski and Pantel (2004), for example, rely on specific patterns constructed manually for each semantic relation between (*similarity, strength, antonymy, enablement* and *temporal happens-before*). They use the web as a corpus in order to estimate the PMI between a pattern and a pair of verbs (a precise measurement cannot be achieved over the web since the probability of a pattern is not precisely known over all the web). A threshold on the value of the PMI (manually fixed) permits thus to determine the pairs of verbs that are related to the relation denoted by the pattern. In the same spirit, Kozareva (2012) is using a weakly supervised approach for the extraction of pairs of verbs that are potentially implied in a *cause-effect* relation. Her method consists in using patterns applied to the web in order to extract pairs and generate new seeds. Do et al. (2011) focus on causal relations and take into account not only verbs but also event denoting nouns. According to this paper, an event is denoted by a predicate with a specific number of arguments and thus the association of the events is the sum of the association between predicates, between predicates and arguments and between arguments. Their association measures are based on PMI and are quite complex. Our results show that their measures do not generalize well to association with all discourse relations. Using Gigaword as a corpus and a reimplementation of Lin et al. (2014) they have extracted discourse relations. An inductive logic programming approach is finally used exploiting the interaction between causal pairs and discourse relations in order to extract causal links. Those papers focus on specific relations with the exception of Chklovski and Pantel (2004) who do not present a systematic evaluation of their results. An important difference of our approach is also to consider predicates and their negation as separate entries.

Finally, we mention the approaches which while focusing on the learning of discourse structures, nonetheless enrich their systems with lexical information. Feng and Hirst (2012) have used HILDA (Hernault et al., 2010) adding more features. A specific family of features represents lexical similarity based on the hierarchical distance in VERBNET and WORDNET. In a similar fashion, Wellner et al. (2006) focus on intra-sentential discourse relations adding lexical information on the features based on measures proposed by Lin (1998) calculated on the British National Corpus. Those approaches use thus only information on lexical similarity without semantically typing this link. The impact of this information seems limited. As far as evaluation is concerned, our method is similar to that followed in Tremper and Frank (2013) for implication relations combining in and out of context evaluation for verbal associations. Their inter-annotator agreement is similar to ours (0.42-0.44 of Kappa) with very different choices : the anno-

tators were supposed to discriminate verbal links between the different possible sub-cases. The pairs of verbs were identified by the system of Lin and Pantel. These authors also present a classification model among the different types of relationships, assuming that two verbs are semantically related.

6 Conclusions

We have presented a knowledge base of triples involving pairs of verbs associated with semantic or discourse relations. We extracted these triples from the large French corpus, frWaC, using discourse connectors as markers of relations between two adjacent clauses containing verbs. We investigated several measures to give the strength of association of a pair of verbs with a relation. We used manual annotations to evaluate our method and select the best measures, and we also studied the coverage of our resource on the discourse annotated corpus Annodis. Our positive results show our resource has the potential to help discourse analysis as well as other semantically oriented tasks.

References

Stergos Afantenos, Nicholas Asher, Farah Benamara, Myriam Bras, Cecile Fabre, Mai Ho-Dac, Anne Le Draoulec, Philippe Muller, Marie-Paul Pery-Woodley, Laurent Prevot, Josette Rebeyrolles, Ludovic Tanguy, Marianne Vergez-Couret, and Laure Vieu. 2012. An empirical resource for discovering cognitive principles of discourse organisation : the ANNODIS corpus. In Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Mehmet Uğur Doğan, Bente Maegaard, Joseph Mariani, Jan Odijk, and Stelios Piperidis, editors, *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)*, Istanbul, Turkey, may. European Language Resources Association (ELRA).

Nicholas Asher and Alex Lascarides. 2003. *Logics of Conversation*. Studies in Natural Language Processing. Cambridge University Press, Cambridge, UK.

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

Marco Baroni, Silvia Bernardini, Adriano Ferraresi, and Eros Zanchetta. 2009. The wacky wide web : a collection of very large linguistically processed web-crawled corpora. *Language resources and evaluation*, 43(3) :209–226.

Chloé Braud and Pascal Denis. 2013. Identification automatique des relations discursives "implicites" à partir de données annotées et de corpus bruts. In *TALN - 20ème conférence du Traitement Automatique du Langage Naturel 2013*, volume 1, pages 104–117, Sables d'Olonne, France, June.

Marie Candito, Benoît Crabbé, and Pascal Denis. 2010. Statistical french dependency parsing : Treebank conversion and first results. In *LREC*.

Jean Carletta. 1996. Assessing agreement on classification tasks : the kappa statistic. *Computational linguistics*, 22(2) :249–254.

Nathanael Chambers and Dan Jurafsky. 2008. Unsupervised Learning of Narrative Event Chains. In *Proceedings of ACL-08 : HLT*, pages 789–797, Columbus, Ohio, June. Association for Computational Linguistics, Morristown, NJ, USA.

Timothy Chklovski and Patrick Pantel. 2004. Verbocean : Mining the web for fine-grained semantic verb relations. In Dekang Lin and Dekai Wu, editors, *Proceedings of EMNLP 2004*, pages 33–40, Barcelona, Spain, July. Association for Computational Linguistics.

P. Denis and B. Sagot. 2012. Coupling an annotated corpus and a lexicon for state-of-the-art pos tagging. *Language Resources and Evaluation*, (46) :721–736.

Quang Do, Yee Seng Chan, and Dan Roth. 2011. Minimally supervised event causality identification. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, pages 294–303, Edinburgh, Scotland, UK., July. Association for Computational Linguistics.

Stefan Evert. 2005. *The statistics of word cooccurrences*. Ph.D. thesis, Stuttgart University.

C. Felbaum. 1998. *Wordnet, an Electronic Lexical Database for English*. Cambridge : MIT Press.

Vanessa Wei Feng and Graeme Hirst. 2012. Text-level discourse parsing with rich linguistic features. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 1 : Long Papers)*, pages 60–68, Jeju Island, Korea, July. Association for Computational Linguistics.

G. Grefenstette. 1994. *Explorations in automatic thesaurus discovery*. Springer.

Chikara Hashimoto, Kentaro Torisawa, Kow Kuroda, Stijn De Saeger, Masaki Murata, and Jun'ichi Kazama. 2009. Large-scale verb entailment acquisition from the Web. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, pages 1172–1181, Singapore, August. Association for Computational Linguistics.

Hugo Hernault, Helmut Prendinger, David A. duVerle, and Mitsuru Ishizuka. 2010. HILDA : A Discourse Parser Using Support Vector Machine Classification. *Dialogue and Discourse*, 1(3) :1–33.

Zornitsa Kozareva. 2012. Cause-effect relation learning. In *Workshop Proceedings of TextGraphs-7 : Graph-based Methods for Natural Language Processing*, pages 39–43, Jeju, Republic of Korea, July. Association for Computational Linguistics.

Dekang Lin and Patrick Pantel. 2002. Concept discovery from text. In *Proceedings of Coling 2002*, pages 1–7. Association for Computational Linguistics.

Ziheng Lin, Hwee Tou Ng, and Min-Yen Kan. 2014. A PDTB-styled end-to-end discourse parser. *Natural Language Engineering*, 20(2) :151–184.

Dekang Lin. 1998. Automatic retrieval and clustering of similar words. In *Proceedings of the 36th ACL and 17th COLING joint conference*, volume 2, pages 768–774, Montreal.

Maofu Liu, Wenjie Li, Mingli Wu, and Qin Lu. 2007. Extractive summarization based on event term clustering. In *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics Companion Volume Proceedings of the Demo and Poster Sessions*, pages 185–188, Prague, Czech Republic, June. Association for Computational Linguistics.

Daniel Marcu and Abdessamad Echihabi. 2002. An Unsupervised Approach to Recognizing Discourse Relations. In *Proceedings of ACL*, pages 368–375.

Seyed Abolghasem Mirroshandel, Alexis Nasr, and Benoît Sagot. 2013. Enforcing subcategorization constraints in a parser using sub-parses recombining. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics : Human Language Technologies*, pages 239–247, Atlanta, Georgia, June. Association for Computational Linguistics.

Philippe Muller, Stergos Afantenos, Pascal Denis, and Nicholas Asher. 2012. Constrained decoding for text-level discourse parsing. In *Proceedings of COLING 2012*, pages 1883–1900, Mumbai, India, December. The COLING 2012 Organizing Committee.

Joakim Nivre, Johan Hall, Jens Nilsson, Atanas Chanev, Gülsen Eryigit, Sandra Kübler, Svetoslav Marinov, and Erwin Marsi. 2007. Maltparser : A language-independent system for data-driven dependency parsing. *Natural Language Engineering*, 13(2) :95–135.

Rashmi Prasad, Nikhil Dinesh, Alan Lee, Eleni Miltsakaki, Livio Robaldo, Aravind Joshi, and Bonnie L. Webber. 2008. The Penn Discourse TreeBank 2.0. In *Proceedings of LREC 2008*.

Charlotte Roze, Laurence Danlos, and Philippe Muller. 2012. Lexconn : A french lexicon of discourse connectives. *Discours*, (10).

Benoît Sagot. 2010. The lefff, a freely available and large-coverage morphological and syntactic lexicon for french. In *7th international conference on Language Resources and Evaluation (LREC 2010)*.

Caroline Sporleder and Alex Lascarides. 2008. Using Automatically Labelled Examples to Classify Rhetorical Relations : An Assessment. *Natural Language Engineering*, 14(3) :369–416, July.

Galina Tremper and Anette Frank. 2013. A discriminative analysis of fine-grained semantic relations including presupposition : Annotation and classification. *Dialogue & Discourse*, 4(2) :282–322.

Naushad UzZaman, Hector Llorens, Leon Derczynski, James Allen, Marc Verhagen, and James Pustejovsky. 2013. Semeval-2013 task 1 : Tempeval-3 : Evaluating time expressions, events, and temporal relations. In *Second Joint Conference on Lexical and Computational Semantics (*SEM), Volume 2 : Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013)*, pages 1–9, Atlanta, Georgia, USA, June. Association for Computational Linguistics.

K. Van Den Eynde and P. Mertens, 2010. *Le dictionnaire de valence : Dicovalence*. Leuven : Université de Leuven. [http ://bach. arts. kuleuven. be/dicovalence/].

Ben Wellner, James Pustejovsky, Catherine Havasi, Anna Rumshisky, and Roser Saurí. 2006. Classification of discourse coherence relations : an exploratory study using multiple knowledge sources. In *Proceedings of the 7th SIGdial Workshop on Discourse and Dialogue*, SigDIAL '06, pages 117–125, Stroudsburg, PA, USA. Association for Computational Linguistics.