

## Extraction non supervisée de relations sémantiques lexicales \*

Juliette Conrath Stergos Afantenos Nicholas Asher Philippe Muller  
IRIT, Université Toulouse & CNRS, Univ. Paul Sabatier, 118 Route de Narbonne, 31062 Toulouse  
{nom.prenom@irit.fr}

**Résumé.** Nous présentons une base de connaissances comportant des triplets de paires de verbes associés avec une relation sémantique/discursive, extraits du corpus français frWaC par une méthode s'appuyant sur la présence d'un connecteur discursif reliant deux verbes. Nous détaillons plusieurs mesures visant à évaluer la pertinence des triplets et la force d'association entre la relation sémantique/discursive et la paire de verbes. L'évaluation intrinsèque est réalisée par rapport à des annotations manuelles. Une évaluation de la couverture de la ressource est également réalisée par rapport au corpus Annodis annoté discursivement. Cette étude produit des résultats prometteurs démontrant l'utilité potentielle de notre ressource pour les tâches d'analyse discursive mais aussi des tâches de nature sémantique.

**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 ressource with respect to the discourse annotated corpus Annodis. Our positive results show the potential impact of our ressource for discourse analysis tasks as well as semantically oriented tasks.

**Mots-clés :** discours, sémantique, sémantique lexicale.

**Keywords:** discourse, semantics, lexical semantics.

### 1 Introduction

Les ressources lexicales relationnelles, c'est-à-dire qui répertorient les liens entre items lexicaux d'un point de vue sémantique, sont essentiellement tournées vers l'équivalence sémantique (synonymie, similarité) dans des thesaurus, éventuellement avec des relations hiérarchiques (hyperonymie ou hyponymie), à l'exemple de la référence Wordnet (Felbaum, 1998), qui inclut aussi des relations de partie à tout. Quand elles incluent des relations plus variées, par exemple dans les thesaurus distributionnels (Grefenstette, 1994), celles-ci ne sont pas typées. Les exceptions sont rares : le lexique sémantique FrameNet (Baker *et al.*, 1998) inclut des relations de causalité ou de précédence temporelle entre items désignant des événements, à l'intérieur de scénarios prototypiques, mais ces relations sont peu nombreuses et relativement négligeables par rapport au contenu de ce lexique ; la base Verbocean (Chklovski & Pantel, 2004) comporte des relations sémantiques de plusieurs types entre verbes transitifs : relations de type causal (*enablement*, par exemple *fight/win*), précédence temporelle (*marry/divorce*), similarité, antonymie, force (*wound/kill*), mais sa couverture est assez faible (environ 4000 paires de verbes dans sa version filtrée). Dans les deux cas les validations sont partielles, et ces travaux semblent avoir été laissés en attente. Les relations lexicales, notamment entre verbes, sont pourtant cruciales pour la compréhension du langage naturel, et sont utilisées par exemple dans la tâche d'inférence textuelle, où il s'agit de trouver les implications entre certains événements (Hashimoto *et al.*, 2009; Tremper & Frank, 2013), dans certaines tâches d'extraction, par exemple de relations temporelles (UzZaman *et al.*, 2013), dans l'analyse discursive en l'absence de marques explicites (Sporleder & Lascarides, 2008), ou bien encore pour le résumé automatique (Liu *et al.*, 2007). Certains travaux se sont consacrés à l'inventaire de tels liens pour des relations spécifiques : par exemple les liens causaux (Do *et al.*, 2011), les liens temporels (Chambers & Jurafsky, 2008), les liens d'implication (*entailment*) (Hashimoto *et al.*, 2009), implication et présupposition (Tremper & Frank, 2013). Le but de notre travail est l'extraction de certaines relations sémantiques qui sont primordiales pour l'analyse discursive. Par analyse discursive nous entendons l'établissement de liens entre énoncés, au-delà de la phrase, comme dans l'exemple (1), où en l'absence de marque explicite (par exemple avec un connecteur comme *donc* ou

\*. Ce travail a bénéficié d'une aide de l'Agence Nationale de la Recherche portant la référence (ANR-12-CORD-0004)

ainsi), on peut déduire une causalité entre les événements par la connaissance de la sémantique des deux verbes mis en évidence.

(1) Le candidat a démontré tout son savoir-faire lors de la dernière épreuve. Le jury a été conquis.

L'analyse discursive est une tâche difficile, y compris pour l'humain. En effet, les relations rhétoriques sont fréquemment implicites et nécessitent des inférences pour être identifiées<sup>1</sup>. Ceci rend l'annotation de corpus fastidieuse et souvent imprécise, et peu de données annotées en relation sont ainsi disponibles à l'heure actuelle. De ce fait, les approches inductives par apprentissage ont un impact réduit, puisque celles-ci nécessitent un très grand nombre de données. De nombreux travaux tentent de pallier ces limitations en élaborant des approches faiblement supervisées, utilisant des données non annotées mais comportant des marques explicites repérables automatiquement pour retrouver des contextes typiques de certaines relations. Ces approches (Sporleder & Lascarides, 2008; Braud & Denis, 2013) s'appuient sur l'hypothèse de régularité des contextes, notamment des associations lexicales, mais de façon indirecte. À l'inverse, certaines approches (Wellner *et al.*, 2006; Feng & Hirst, 2012) tentent d'enrichir les modèles d'apprentissage avec des relations lexicales fines, du type de celles mentionnées plus haut, mais se heurtent à la faible couverture des ressources existantes.

Lister explicitement toutes les associations possibles de deux verbes dans cette perspective semble difficilement réalisable manuellement, et nous présentons une approche automatique, inspirée du projet Verbocean (Chklovski & Pantel, 2004), pour constituer une base lexicale large associant des verbes avec un typage d'ordre sémantique. Si notre objectif final est d'aider à la prédiction de relations discursives (rhétoriques), nous pensons utile de constituer cette ressource lexicale comme un intermédiaire intéressant d'autres tâches de nature sémantique. L'essentiel de la méthode est de recenser les paires de verbes dans des clauses adjacentes dans un grand corpus. Ces clauses sont souvent liées par des adverbiaux ou plus généralement des connecteurs discursifs marquant une ou plusieurs relations discursives ; ces marqueurs suggèrent une relation discursive qui est enregistrée en association avec la paire. L'idée est de récupérer des paires de verbes qui sont souvent marquées avec une relation discursive (ou temporelle) et d'en déduire que cette paire de verbes peut suggérer une telle relation même en l'absence de marqueur. Ceci suppose que la relation repose non seulement sur le connecteur, mais également sur la paire de verbes employée : nous faisons ainsi l'hypothèse de redondance partielle du marqueur. Cette hypothèse a été précédemment discutée dans la littérature, notamment par (Sporleder & Lascarides, 2008) et (Braud, 2011), qui tendent à la soutenir.

La suite de l'article est organisée de la façon suivante. Nous détaillons d'abord la base de connaissances que nous avons construite (section 2), puis nous présentons ensuite les méthodes utilisées pour isoler des paires de verbes susceptibles d'apporter des informations discursives ou temporelles (section 3). Une troisième section décrit nos méthodes d'évaluation (section 4) et une quatrième fait une comparaison avec d'autres approches (section 5).

## 2 Explorer les relations entre verbes en corpus

La base de connaissances de relations verbales<sup>2</sup> a été construite à partir du corpus frWaC, qui fait partie de l'ensemble de corpora WaCKy (Baroni *et al.*, 2009). Celui-ci a été collecté sur le Web dans le domaine `.fr`, et contient environ 1.6 milliards de mots. Ce corpus a d'abord été parsé syntaxiquement grâce à la chaîne d'analyse de texte BONSAI<sup>3</sup> : étiquetage morpho-syntaxique par l'outil MELT (Denis & Sagot, 2012), puis analyse syntaxique en dépendances via une adaptation française du MaltParser (Nivre *et al.*, 2007). Le format de sortie est de type CONLL.

L'objectif est de rechercher des paires de verbes liés par une relation marquée explicitement par un connecteur dans le corpus. Les relations considérées sont des relations typiques en analyse discursive, éventuellement regroupée en groupes cohérents. Le corpus anglais du Penn Discourse TreeBank, le PDTB, (Prasad *et al.*, 2008) définit ainsi une hiérarchie d'une trentaine de relations comportant un niveau supérieur de quatre groupes : un groupe de relations causales (*contingency*), un groupe de relations temporelles, un groupe de "comparaison" (essentiellement des liens contrastifs), et un groupe d'"expansion" (essentiellement des types d'élaboration ou de continuation de discours). Afin de repérer les relations discursives marquées explicitement, on dispose en français du lexique LEXCONN (Roze *et al.*, 2012)<sup>4</sup>, construit manuellement. Celui-ci rassemble 358 connecteurs du discours et comprend leurs catégories syntaxiques et les relations discursives associées, relations proches de celles de la SDRT (Asher & Lascarides, 2003). Certains connecteurs ont un usage ambigu, ils peuvent être associés à plusieurs relations. Dans un premier temps, dans un but de simplification, nous

1. Le corpus anglais PDTB par exemple comporte 52% de relations non marquées (Prasad *et al.*, 2008).

2. Disponible sous forme de base de données SQLite à [https://dl.dropboxusercontent.com/u/78938139/v2r\\_db](https://dl.dropboxusercontent.com/u/78938139/v2r_db)

3. Disponible à [http://alpage.inria.fr/statgram/frdep/fr\\_stat\\_dep\\_parsing.html](http://alpage.inria.fr/statgram/frdep/fr_stat_dep_parsing.html), cf aussi (Candito *et al.*, 2010)

4. Disponible librement : <https://gforge.inria.fr/frs/download.php/31052/lexconn.tar.gz>.

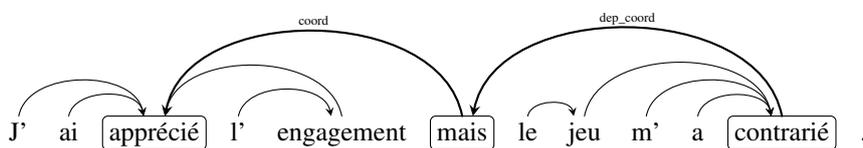
avons choisi de ne conserver que les connecteurs non ambigus, au nombre de 263. Par la suite, une désambiguïssation pourra être opérée afin de permettre la prise en compte de tous les connecteurs. LEXCONN distingue une vingtaine de relations, et comme pour le PDTB, nous avons constitué des regroupements significatifs<sup>5</sup> : les relations d'explication (*parce que*) et de résultat (*ainsi*) forment le groupe causal, les relations d'organisation temporelle (*puis, après que*) ont été regroupées en un groupe de relations de narration. Les autres relations considérées sont des relations structurelles de contraste (*mais*), continuation (*et, encore*), arrière-plan (*alors que*), localisation temporelle (*quand, pendant que*), détachement (*de toutes façons*), élaboration (*en particulier*), alternation (*ou*), commentaire (*au fait*), reformulation (*du moins*), évidence (*effectivement*).

Un parcours du corpus parsé syntaxiquement est donc réalisé à la recherche de ces relations. Lorsqu'un connecteur est rencontré (après vérification de sa catégorie syntaxique), si celui-ci se trouve suffisamment proche de la racine de la phrase, une relation interphrastique est recherchée. Le premier verbe de la paire correspond alors au dernier verbe de la phrase précédente pour le cas des connecteurs de narration, ou au verbe principal de celle-ci pour toutes les autres relations. Le second verbe de la paire est recherché dans une fenêtre de deux liens de dépendances après le connecteur. Si le connecteur n'est pas suffisamment proche de la racine, une relation intraphrastique est recherchée. Pour cela, les deux verbes de la paire sont recherchés au sein de la même phrase, de part et d'autre du connecteur dans une fenêtre de deux liens de dépendances.

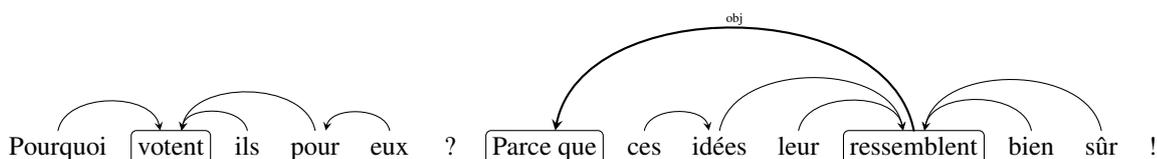
Dans le cas où deux lemmes verbaux sont effectivement identifiés, le contexte est examiné afin de mieux caractériser leur usage et d'affiner nos résultats. D'abord, si un verbe détecté est un modal ou un verbe support, la focalisation est reportée sur le verbe supporté par celui-ci, s'il existe, tout en mémorisant la présence du verbe support (qui ne constitue pas une distinction). La présence ou absence de négation et de particule réflexive constituent des critères de distinction entre les verbes (*comprendre* et *ne pas comprendre*, *agir* et *s'agir* sont des entrées distinctes). Par ailleurs, afin de distinguer les différents sens des verbes, deux types de traitement sont effectués. Une recherche d'un usage idiomatique de préposition est réalisée grâce à la ressource Dicovalence (Van Den Eynde & Mertens, 2010), qui répertorie les cadres de valence de plus de 3700 verbes simples du français (exemple : *tenir de* et *tenir à*). De plus la ressource Lefff (Lexique des Formes Fléchies du Français) (Sagot, 2010) permet de repérer les locutions verbales (exemples : *prendre garde, faire référence*). D'autres informations sont également mémorisées mais ne sont pas distinctives : temps du verbe, voix passive ou active.

Les deux exemples suivants illustrent cette méthode d'extraction (le schéma de dépendance est présenté, avec les liens utilisés pour l'extraction représentés en gras).

Exemple de relation intraphrastique :



Exemple de relation interphrastique :



Une fois la liste des paires associées à un connecteur obtenue, une agrégation de ces résultats est effectuée afin de les regrouper en types de triplets distincts (verbe 1, verbe 2, relation). Étant donné que seuls les connecteurs non ambigus ont été conservés, l'obtention de la relation associée est directe. À chaque triplet sont associés le nombre d'occurrences intraphrastiques, interphrastiques, et le nombre total d'occurrences. Les autres données récoltées mentionnées plus haut (temps, verbe support...) sont également conservées dans une table annexe.

Par cette méthode, plus de 2 millions d'occurrences de triplets, dont près de 95% intraphrastiques<sup>6</sup>, ont été obtenues

5. Pour illustrer chaque relation nous donnons des exemples de marqueurs explicites entre parenthèses, potentiellement ambigus.

6. La faible proportion d'occurrences interphrastiques provient du schéma prudent choisi pour le repérage de ces occurrences, considérant uniquement les connecteurs présents en début de la seconde phrase. Nous avons en effet jugé le risque de produire du bruit en élargissant les schémas possibles trop important.

dans le corpus. Ces occurrences ont été regroupées en plus d'1 million de types de triplets distincts dans la base de connaissances. Parmi ces triplets, 4,5% présentent 5 occurrences ou plus.

Relation	Distribution
contraste	50,104%
cause	33,108%
continuation	8,243%
narration	6,362%
arrière-plan	1,853%
localisation temporelle	0.177%
détachement	0.149%
élaboration	0.002%
alternation	0.002%

TABLE 1 – La distribution des relations dans la base ; les relations commentaire, reformulation et évidence ont une fréquence d'apparition presque nulle.

La table 1 résume la distribution des triplets par relation. Nous pouvons noter que les relations de contraste et de cause sont largement majoritaires dans la base. Ceci ne signifie pas forcément que ce sont les plus présentes dans le corpus, mais plutôt qu'elles sont le plus fréquemment marquées par les connecteurs considérés. En prenant comme point de comparaison la ressource Annodis (Afantenos *et al.*, 2012), un corpus annoté manuellement en relations de discours, nous pouvons ainsi remarquer que les relations de continuation et d'élaboration sont bien plus fréquentes que dans nos annotations automatiques. Ceci signifie que ces relations reposent probablement moins sur la présence d'une marque explicite telle qu'un connecteur.

### 3 Mesurer l'association sémantique des paires de verbes

La section précédente présentait les données collectées sur les paires de verbes reliées par les marqueurs choisis, nous présentons maintenant les mesures testées pour classer la force d'association des paires de verbes. Nous nous sommes inspirés des mesures classiques d'association lexicale, issues de l'étude des cooccurrences, en les adaptant au contexte, et avons ajouté certaines mesures utilisées sur des relations spécifiques, comme les mesures de Do *et al.* (2011) pour détecter les liens de causalité entre verbes.

Les mesures d'association lexicale utilisées dans la recherche de cooccurrences servent à repérer des associations significatives, une fois que l'on tient compte de la fréquence des items reliés. Elles sont un composant essentiel des approches distributionnelles de la sémantique et dans la construction d'espaces vectoriels de mots. Nous avons retenu une mesure simple, la PMI (*pointwise mutual information*) et ses variantes, *locale*, *normalisée*, *pondérée*, sensées atténuer les biais de la mesure originale (Evert, 2005). Le principe de la PMI est d'estimer si l'apparition simultanée de deux items est supérieure à la probabilité d'apparition *a priori* des deux items indépendamment. Nous avons appliqué cette mesure à des triplets constitués d'une paire de verbes avec une relation sémantique/discursive, parce que ce qui nous intéresse est de voir si la probabilité des deux items *avec* une relation sémantique particulière est supérieure à la probabilité d'apparition *a priori* des trois items indépendamment. Pour toutes nos mesures, nous considérons en fait l'événement consistant en l'apparition de deux items lexicaux dans une certaine relation indiquée par un marqueur explicite. Ceci est semblable aux approches syntaxiques en sémantique distributionnelle, qui pondèrent les associations d'items lexicaux dans une certaine relation syntaxique (comme nom-sujet-verbe, ou verbe-objet-nom).

$$PMI = \log\left(\frac{P(V_1, V_2, R)}{P(V_1) \times P(V_2) \times P(R)}\right)$$

En cas de cooccurrence complète des trois items, nous avons :

$$P(V_1) = P(V_2) = P(R) = P(V_1, V_2, R), \text{ et } PMI = -2 \log(P(V_1, V_2, R)).$$

Ainsi, la PMI normalisée est définie comme suit :

$$PMI_{normalisée} = \frac{PMI}{-2 \log(P(V_1, V_2, R))}$$

Notons ainsi que cette mesure est comprise entre -1 et 1, approchant -1 lorsque les items n'apparaissent jamais ensemble, prenant la valeur 0 en cas d'indépendance, et la valeur 1 en cas de cooccurrence complète.

La PMI pondérée proposée par Lin & Pantel (2002) est censée pallier le biais de la PMI pour les triplets peu fréquents :

$$PMI\_pondérée = discount \times PMI$$

$$discount = \frac{P(V_1, V_2, R)}{P(V_1, V_2, R) + 1} \times \frac{\min[\sum_i (P(V_i, V_2, R)), \sum_i (P(V_1, V_i, R)), \sum_i (P(V_1, V_2, R_i))]}{\min[\sum_i (P(V_i, V_2, R)), \sum_i (P(V_1, V_i, R)), \sum_i (P(V_1, V_2, R_i))] + 1}$$

La PMI locale quant à elle permet de prendre en compte la fréquence absolue d'occurrence du triplet :

$$PMI\_locale = F(V_1, V_2, R) \times \log\left(\frac{P(V_1, V_2, R)}{P(V_1) \times P(V_2) \times P(R)}\right) = F(V_1, V_2, R) \times PMI$$

Nous nous sommes également inspirés d'une mesure de (Mirroshandel *et al.*, 2013), initialement définie pour mesurer la précision de cadres de sous-catégorisation, pour définir la mesure de spécificité :

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

Do *et al.* (2011) donnent une mesure complexe pour l'apport de deux prédicats qui supportent une relation causale, dont nous nous sommes inspirés dans la mesure suivante :

$$U_{do}(V_1, V_2, R) = PMI(V_1, V_2, R) \times \max\{u_{V_1}, u_{V_2}, u_R\}$$

$$\text{où : } u_{V_1} = \frac{P(V_1, V_2, R)}{\max_i (P(V_i, V_2, R)) - P(V_1, V_2, R) + \varepsilon}, \quad u_{V_2} = \frac{P(V_1, V_2, R)}{\max_i (P(V_1, V_i, R)) - P(V_1, V_2, R) + \varepsilon}$$

$$\text{et } u_R = \frac{P(V_1, V_2, R)}{\max_i (P(V_1, V_2, R_i)) - P(V_1, V_2, R) + \varepsilon}.$$

Notons que la mesure initiale de (Do *et al.*, 2011) est également fonction de l'IDF (inverse document frequency), qui mesure la fréquence interdocument des deux verbes, et de la distance entre les deux instances. Ces deux derniers facteurs ne sont pas applicables à nos triplets, et ont donc été ignorés.

Nous avons également défini une mesure permettant d'évaluer l'apport de chaque composant du triplet à son informativité, similaire à la spécificité décrite ci-dessus.

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

$$\text{Avec : } w_{V_1} = \frac{P(V_1, V_2, R)}{\max_i (P(V_i, V_2, R))}, \quad w_{V_2} = \frac{P(V_1, V_2, R)}{\max_i (P(V_1, V_i, R))}, \quad \text{et } w_R = \frac{P(V_1, V_2, R)}{\max_i (P(V_1, V_2, R_i))}.$$

## 4 Évaluation des relations

Pour évaluer l'intérêt des paires de verbes extraites, nous avons procédé à plusieurs évaluations qui se veulent complémentaires. D'abord nous voulons une évaluation intrinsèque du lien entre les verbes, dans la perspective de valider la base comme une ressource sémantique, qui peut servir à des tâches différentes. Celle-ci est présentée ci-dessous (section 4.1). Nous présentons ensuite un début de validation extrinsèque, en étudiant l'impact potentiel de la ressource sur une tâche spécifique, à savoir la prédiction de relations discursives en l'absence de marque explicite (section 4.2).

### 4.1 Evaluation intrinsèque

Pour évaluer intrinsèquement les liens extraits, nous avons dans un premier temps étudié la possibilité d'attribuer fiablement un lien sémantique à une paire de verbes de façon "inhérente", c'est-à-dire hors de tout contexte Par exemple pour la

cause, est-il possible de juger qu’il y a une causalité “typique” entre les verbes *pousser* et *tomber*, dans des scénarios où ils partagent des arguments (sujet, objet, ...), ces scénarios étant laissés à l’appréciation du juge (section 4.1.1). Dans un deuxième temps, nous avons sélectionné quelques paires de verbes et une centaine de contextes dans lesquels ces paires apparaissent ensemble dans le corpus d’origine, pour juger du lien sémantique en contexte (section 4.1.2). Dans les deux cas, nous avons restreint l’étude à trois groupes de relations, causales, contrastives et narratives. Ce sont les plus couramment marquées dans le corpus, et elles constituent des cas assez différents de lien, en ayant une composante sémantique qui paraît significative (à l’inverse de la relation, très marquée également, de continuation).

#### 4.1.1 Evaluation hors contexte des paires

Pour le jugement sur des liens hors contexte, nous avons suivi le protocole suivant : un des auteurs a choisi 100 paires de verbes avec des proportions similaires de paires présentant des bons et mauvais scores pour la relation choisie et selon les mesures choisies. Ensuite, les trois autres auteurs ont dû juger pour chacune des 300 paires si elle pouvait ou non être reliée avec la relation considérée, sans connaître l’origine des paires. La table résume les accords inter-annotateurs, estimés avec le kappa de Cohen (Carletta, 1996). Il est apparu assez vite que la tâche était très difficile voire infaisable pour la causalité, difficile pour la narration, et moyennement difficile pour le contraste, si l’on juge classiquement qu’un kappa autour de 0.6 est acceptable, surtout pour un jugement de nature sémantique.

Nous avons donc décidé d’ignorer ces jugements pour la cause et la narration, et avons gardé les jugements sur le contraste, après adjudication entre les trois annotateurs. Pour évaluer les mesures d’association choisies, nous avons testé statistiquement si elles discriminaient entre les deux groupes de paires de verbes (celles jugées positivement et négativement par les annotateurs). La table 3 résume ces tests, où l’on voit que toutes les mesures discriminent statistiquement les deux groupes, sauf les comptages bruts de cooccurrence.

Annotateurs	Cause	Contraste	Narration
1/2	0.16	0.55	0.43
1/3	0.22	0.57	0.46
2/3	0.13	0.56	0.37
kappa moyen	0.17	0.56	0.42

TABLE 2 – Accords inter-annotateurs pour annotations hors contexte : kappa par paires d’annotateurs et moyenne des kappas.

Mesure	valeur de p
spécificité	2.5e-11
U_do	2.9e-11
PMI_normalisée	1.28e-10
PMI_pondérée	1.96e-10
PMI	1.86e-10
W_combinée	4.93e-10
PMI_locale	4.95e-08
comptage d’occurrences inter-phrastiques	0.000904
comptage d’occurrences intra-phrastiques	0.0721
comptage d’occurrences brut	0.116

TABLE 3 – Tests de MannWhitney-U pour estimer la différence des mesures sur les groupes de paires de verbes contrastives ou non. Les mesures sont triées par valeur de p croissante, les valeurs dans la deuxième partie du tableau étant non significatives.

#### 4.1.2 Evaluation en contexte

Pour une évaluation plus précise des liens sémantiques, nous avons aussi considéré des jugements d’association en contexte : en plus de faciliter le jugement, l’idée est aussi que le caractère typique de la relation entre deux verbes peut être

estimée par la proportion de contextes où ils apparaissent ensemble avec le lien que l'on considère. On peut ensuite observer si cette proportion est corrélée avec les mesures d'association présentées précédemment. L'écueil de cette méthode est son coût en annotation : si l'on veut évaluer un lien entre deux verbes spécifiques il faut déjà un certain nombre de contextes pour appuyer le jugement, et il faut répéter cette évaluation sur un nombre suffisant de paires de verbes. De plus, il faut avoir un échantillon de paires qui couvre suffisamment de valeurs différentes pour observer des corrélations significatives, alors qu'on ne peut préjuger des valeurs attribuées par l'annotation humaine. Nous avons finalement choisi 40 contextes exemples pour chacune des 15 paires de verbes sélectionnées (5 pour chaque relation : cause, narration, contraste) ; les paires sélectionnées le sont selon des scores échelonnés de PMI\_normalisée, et encore une fois par l'un des auteurs indépendamment des trois autres, qui ont réalisé l'annotation séparément, avant de procéder à une adjudication des 600 contextes pour la référence. Pré-adjudication, l'accord brut sur les décisions est de 78% en moyenne, pour un kappa moyen de 0,46 et un kappa maximum de 0,49. Ces valeurs semblent faibles, ce qui souligne la difficulté de la tâche. Le résultat de cette annotation est présenté table 4.

Verbe 1	Verbe 2	relation	association/humain
inviter	souhaiter	causale	12.8%
promettre	élire	causale	25.6%
aimer	trouver	causale	38.5%
bénéficier	créer	causale	51.3%
aider	gagner	causale	53.8%
proposer	refuser	contraste	59.0%
augmenter	diminuer	contraste	64.1%
tenter	échouer	contraste	64.1%
gagner	perdre	contraste	71.8%
autoriser	interdire	contraste	74.4%
parler	réfléchir	narration	42.5%
acheter	essayer	narration	70.0%
atteindre	traverser	narration	77.5%
commencer	finir	narration	80.0%
envoyer	transmettre	narration	82.5%

TABLE 4 – La liste des paires de verbes évaluées manuellement en contexte, avec la relation à juger et le ratio d'association résultant de l'adjudication humaine

Nous avons ensuite mesuré la corrélation entre cet indice d'association obtenu à partir de l'annotation humaine, et les mesures d'association présentées plus haut. Nous indiquons ici deux corrélations séparées : une première sur l'ensemble des données annotées, et une seconde sur le sous-ensemble des contextes ne comportant pas de marqueur de la relation considérée (les contextes implicites). Cette dernière mesure est importante pour quantifier l'apport effectif de la méthode suivie ici, et éviter la tautologie qui consisterait à trouver des liens marqués explicitement en corpus pour identifier les mêmes liens marqués dans des contextes particuliers. De fait, les contextes sans marques explicites sont les seuls à n'être pas intervenu dans le calcul des mesures d'association.

	Corrélation globale	Corrélation sur instances implicites
PMI_normalisée	0.749	0.806
spécificité	0.747	0.760
W_combinée	0.720	0.738
PMI_pondérée	0.716	0.761
PMI	0.709	0.756
PMI_locale	0.434	0.553
U_do	0.376	0.499
frequence brute	0.170	0.242

TABLE 5 – Les corrélations de Pearson pour les 15 paires considérées et pour les mesures présentées à la section 3, par ordre décroissant.

Ceci posé, on peut observer que les mesures d'information mutuelle sont bien corrélées, même la PMI simple, et que la mesure  $W_{\text{combinée}}$  que nous proposons section 3 est également utile. Nous avons également observé les résultats pour chaque relation séparément (résultats non détaillés ici), avec prudence à cause du faible nombre de points (5 par relation), et avons pu noter de grandes variations dans le comportement des mesures sur chaque relation. Notons ainsi que la mesure  $U_{\text{do}}$ , initialement formulée pour les relations causales, ne se généralise pas bien sur l'ensemble des relations ni sur les autres relations, par contre son bon fonctionnement a pu être vérifié pour les relations causales. Par ailleurs, la mesure de  $PMI_{\text{locale}}$  fonctionne très bien pour les relations de narration et de cause.

Ces résultats nous ont permis d'identifier les trois meilleures mesures : la  $PMI_{\text{normalisée}}$ , la spécificité et  $W_{\text{combinée}}$ . Nous avons observé que ces deux dernières assignent leur valeur maximale à de multiples paires. Nous avons alors imposé un ordre lexicographique en utilisant la  $PMI_{\text{normalisée}}$  pour départager les meilleures paires. La table 6 présente ainsi les meilleures paires obtenues pour les relations de narration, continuation, cause et contraste.

Verbe 1	Verbe 2	Relation
abandonner	mener	arrière-plan
ne pas s'arrêter	rouler	narration
donner satisfaction sur	réélire	continuation
emporter	ne pas cesser	élaboration
emprunter	assurer	cause
ne pas manquer	prolonger	détachement
ratifier	trembler	arrière-plan
avoir honte	faire pitié	cause
avoir droit	cotiser pour	loc. temp.
ne pas représenter	stéréotyper	loc. temp.

TABLE 6 – Listes des 10 meilleures paires de la base selon notre ordre lexicographique.

## 4.2 Evaluation extrinsèque

L'objectif principal de notre travail est de constituer une ressource de relations sémantiques, dont la principale application visée est l'aide à la prédiction de relations rhétoriques. Afin d'évaluer la performance de notre base de triplets dans cette optique, nous avons pour perspective de l'utiliser comme traits additionnels dans un modèle de prédiction de relations. Au préalable, nous avons cherché à évaluer l'impact potentiel de notre ressource sur la tâche de prédiction, en étudiant sa couverture par rapport à un corpus en français annoté en relations de discours, le corpus Annodis (Afantenos *et al.*, 2012). Pour espérer améliorer les modèles existants, il faut en effet qu'une partie significative des relations à prédire implique des paires de verbes présentes dans la base constituée. Un indicateur fort du succès potentiel de la tâche est aussi la part de relations entre deux segments contenant deux verbes existant dans la base, et dont le lien majoritaire fait partie du groupe concerné par la relation, par exemple pour une relation d'explication, le fait que les deux verbes soient reliés par un lien causal. Ceci n'est intéressant qu'à la condition que l'instance en question ne présente pas déjà un marquage direct de la relation via un connecteur de discours, c'est-à-dire que la relation soit implicite. Par ailleurs, toujours dans le cas des liens implicites, il est intéressant d'avoir l'information que deux verbes sont dans un certain rapport sémantique, même si celui-ci ne correspond pas directement à un groupe lié à la relation recherchée : un lien causal potentiel entre deux événements peut informer au moins qu'une succession temporelle est pertinente pour considérer le lien entre deux unités de discours.

Pour mener cette étude, nous nous sommes reposés là encore sur la base de marqueurs Lexconn. En première approximation, nous avons considéré qu'une instance de relation entre deux segments de discours était explicite quand un connecteur de Lexconn était présent dans un des deux segments, et qu'un de ses sens recensés était celui de la relation présente dans le corpus. Cela peut au pire surestimer le nombre d'exemples explicites, et assure que les exemples implicites considérés le sont effectivement (à quelques erreurs de détection de marquage près). Pour simplifier nous n'avons considéré que les liens entre unités discursives simples, Annodis comportant également des relations entre ensembles de segments simples (segments dits "complexes").

La table 7 présente les résultats de couverture, pour les relations principales. Il faut noter qu'un petit nombre d'instances du corpus sont concernées (400 sur environ 2000 relations entre segments simples), les autres n'impliquant souvent qu'un

verbe, et certaines étant oubliées à cause d'erreur d'étiquetage ou de détection. Là encore ces chiffres sont à prendre comme des estimations conservatrices.

	global	narration	cause	contrast	elab.	cont.	AR	autres
paires dans annodis	407	72	65	40	97	93	24	16
paires annodis ∈ vpdb	66.8	65.3	67.7	72.5	69.1	60.2	79.2	62.5
triplets annodis ∈ vpdb	33.7	34.7	49.2	65.0	0.0	19.4	8.3	0.0
triplets annodis marqués dans l'instance	20.4	29.2	21.5	62.5	1.0	5.4	12.5	0.0
paires annodis implicites (pas de connecteur/ou autre connecteur)	79.6	70.8	78.5	37.5	99.0	94.6	87.5	100.0
triplets annodis implicites ∈ vpdb (avec relation correcte)	24.6	23.6	40.0	27.5	0.0	18.3	8.3	0.0
paires annodis implicites ∈ vpdb (toutes relations)	53.1	48.6	50.8	27.5	68.0	57.0	70.8	62.5
verbes absents de vpdb	0.4	1.0	0.0	1.4	0.0	0.0	0.0	0.0

TABLE 7 – Couverture des paires de verbes dans la base (vpdb) par rapport aux instances du corpus Annodis impliquant deux verbes. Paire = paire de verbes dans des segments reliés par une relation rhétorique, Triplet=la paire de verbes associée à la relation rhétorique dans une instance du corpus. A part la première ligne, tous les chiffres sont en pourcentage. AR = arrière-plan, cont=continuation, elab=élaboration.

Nous avons listé plusieurs types d'information : la présence dans la base des paires de verbes des instances du corpus discursif, la présence des paires de verbes associés à la relation qui les lie dans l'instance d'Annodis considérée<sup>7</sup>, c'est-à-dire la présence du triplet (verbe1,verbe2,relation) dans la base, et la restriction de ces statistiques aux contextes où il y a présence ou absence d'un marqueur explicite d'une relation.

Nous pouvons voir que presque tous les verbes présents dans le corpus discursif sont recensés dans la base dans au moins une paire (les deux exceptions sont un verbe non lemmatisé par le tagueur, et une locution verbale), mais que les paires recensées dans la base ne couvrent que partiellement les paires apparaissant dans Annodis. En effet entre 60 et 70% des instances associent des verbes présents dans la base (selon la relation), et un peu moins si l'on considère les instances implicites (environ 50% en moyenne), sauf la relation de contraste, fortement marquée dans Annodis. Il est très encourageant de noter qu'une forte proportion de contextes sans marquage contiennent des paires de verbes qui sont collectées dans un contexte marqué, même pour des relations peu marquées comme élaboration ou continuation, et qui plus est une bonne part de ces contextes (plus de la moitié) sont correctement associés à la bonne relation dans la base (éventuellement parmi d'autres). L'hypothèse de redondance partielle des connecteurs semble pouvoir être utile quand on considère le corpus dans son ensemble pour isoler des associations verbales pertinentes pour le discours. Tout ceci demande à l'évidence d'être poursuivi en intégrant ces informations à un étiquetteur de relations discursives, mais semble prometteur dans cette perspective.

## 5 Travaux reliés

Nous pouvons distinguer deux différents types de travaux reliés à notre approche. Pour le premier groupe, l'idée fondamentale est de pallier le manque de données annotées en utilisant une approche faiblement supervisée, exploitant la présence de marqueurs explicites dans un grand corpus non-annoté. Chaque paire d'unités discursives élémentaires est ainsi annotée automatiquement avec la relation discursive correspondant au marqueur (les marqueurs sont souvent filtrés par rapport à leurs usages non-discursifs). Ensuite, ces marqueurs sont éliminés du corpus afin d'empêcher les modèles de se baser sur cet indice, créant ainsi artificiellement des relations implicites. L'article pionnier de cette approche est celui de (Marcu & Echiabi, 2002). Les relations utilisées dans cet article correspondent à un niveau de granularité plus grossier par rapport aux relations typiquement utilisées en RST (Mann & Thompson, 1988), obtenant néanmoins des scores assez bas. La même approche a été aussi poursuivie par Sporleder & Lascarides (2008) obtenant des résultats à peine au-dessus du hasard comme l'ont montré Braud & Denis (2013). Ces derniers ont ainsi observé les performances relativement faibles de cette méthode de prédiction des relations implicites avec des données « artificielles » (relation explicite rendue

7. Chaque paire de verbes extraite d'annodis est présente une seule fois.

artificiellement implicite par suppression du marqueur) par rapport aux résultats obtenus avec des données « naturelles » (relation implicite annotée par un humain). Ils ont alors proposé une méthode consistant à combiner ces deux types de données soit au niveau du jeu de données soit directement au niveau de l'algorithme d'apprentissage, obtenant ainsi une amélioration significative sur le corpus ANNODIS. Notre approche est différente en ce qu'elle veut isoler explicitement les liens entre paires de verbes, pour élargir l'usage de cette information à d'autres tâches. Dans une prochaine étape nous chercherons cependant à mesurer l'apport de notre ressource pour cette tâche de prédiction de relations par rapport aux approches précédentes, permettant ainsi une évaluation extrinsèque.

Un deuxième groupe de travaux vise à identifier les relations discursives (implicites ou non) en se focalisant sur l'utilisation des relations lexicales fines comme un autre indice pendant la phase d'apprentissage. La plupart des travaux se concentrent principalement sur les relations lexicales entre deux verbes. Chklovski & Pantel (2004) par exemple, se sont appuyés sur des patrons spécifiques construits manuellement pour chaque relation sémantique parmi (*similarity*, *strength*, *antonymy*, *enablement* et *temporal happens-before*). Ensuite, le Web a servi de corpus afin d'estimer la PMI entre deux verbes et un patron (un calcul précis de ne peut pas être réalisable puisque la probabilité d'un verbe ou un patron sur tout le web ne peut être connue précisément). Un seuil (estimé manuellement) sur les valeurs de PMI a ainsi permis de déterminer les paires de verbes considérées comme liées par la relation indiquée par le patron. Dans le même esprit, Kozareva (2012) s'est basée sur une approche faiblement supervisée pour réaliser l'extraction de paires de verbes potentiellement impliqués dans une relation *cause-effet*. La méthode consiste à utiliser des patrons appliqués sur le Web pour extraire des paires et générer de nouvelles graines. Des travaux similaires ont été réalisés par Do *et al.* (2011), prenant cependant en compte non seulement les verbes mais aussi les noms dénotant un événement. Ils se sont concentrés sur les relations causales, utilisant les marqueurs discursifs comme indice. Selon leurs travaux, un événement est un prédicat avec un certains nombre d'arguments et donc l'association d'événements est la somme d'associations entre prédicats, entre prédicats et arguments et entre arguments. Toutes leurs mesures sont basées sur la PMI corrigée pour certains cas (paires trop fréquentes, distance textuelle entre les prédicats d'une paire, fréquence des prédicats). À l'aide du Gigaword comme corpus et d'une réimplémentation de (Lin *et al.*, 2014), ils ont alors extrait les relations discursives. Un système de programmation logique inductive est finalement utilisé, exploitant les interactions entre paires causales et relations discursives afin d'extraire les liens causaux. Ces travaux se concentrent donc sur des relations particulières, à l'exception de Chklovski & Pantel (2004), qui ne présentent pas d'évaluation systématique de leurs résultats.

Enfin, il faut mentionner les travaux qui se soucient directement de l'apprentissage des structures discursives mais qui enrichissent leur système en ajoutant de l'information lexicale. Feng & Hirst (2012) ont utilisé HILDA (Hernault *et al.*, 2010), y ajoutant d'avantage de traits. Une famille de traits représente la similarité lexicale fondée sur les distance dans les hiérarchies VERBNET et WORDNET. D'une façon similaire, Wellner *et al.* (2006) se sont focalisés sur les relations discursives intraphrastiques et ont ajouté de l'information lexicale dans des traits basés sur les mesures proposées par Lin (1998) et calculées sur le British National Corpus. Ces approches n'utilisent donc que des liens lexicaux de similarité, sans typage sémantique de ce lien, et l'impact de cette information seule semble limitée. Du point de vue de l'évaluation, notre méthode est assez proche de celle suivie dans les travaux sur la relation d'implication dans (Tremper & Frank, 2013), combinant évaluation hors et en contexte d'associations verbales. Les accords inter-annotateurs sont similaires aux nôtres (0.42-0.44 de Kappa), avec des choix légèrement différents : les annotateurs étaient censés discriminer le lien verbal entre les différents sous-cas possibles. Les paires de verbes étaient repérées par le système de Lin et Pantel. Ces auteurs présentent également un modèle de classification parmi les différents types de relations, en supposant donné le fait que deux verbes sont liés sémantiquement.

## 6 Conclusion

Nous avons présenté ici une base de connaissances comportant des triplets de paires de verbes associés avec une relation sémantique/discursive, extraits du corpus français frWaC par une méthode s'appuyant sur la présence d'un connecteur discursif reliant deux verbes. Nous avons détaillé plusieurs mesures visant à évaluer la pertinence des triplets et la force d'association entre la relation sémantique/discursive et la paire de verbes. Des évaluations par annotation manuelle nous ont permis de valider notre approche et de sélectionner les meilleures mesures. Nous avons également réalisé une étude de la couverture de la ressource par rapport aux triplets annotés manuellement du corpus Annodis. Ceci nous a permis de vérifier qu'un grand nombre de triplets implicites dans Annodis sont présents dans notre base de connaissances. Ces résultats positifs encouragent la poursuite de nos travaux dans la perspective d'utiliser cette ressource pour améliorer les méthodes d'analyse du discours, mais aussi pour des tâches de nature sémantique.

## Références

- AFANTENOS S., ASHER N., BENAMARA F., BRAS M., FABRE C., HO-DAC M., DRAOULEC A. L., MULLER P., PERY-WOODLEY M.-P., PREVOT L., REBEYROLLES J., TANGUY L., VERGEZ-COURET M. & VIEU L. (2012). An empirical resource for discovering cognitive principles of discourse organisation : the ANNODIS corpus. In N. CALZOLARI, K. CHOUKRI, T. DECLERCK, M. U. DOĞAN, B. MAEGAARD, J. MARIANI, J. ODIJK & S. PIPERIDIS, Eds., *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)*, Istanbul, Turkey : European Language Resources Association (ELRA).
- ASHER N. & LASCARIDES A. (2003). *Logics of Conversation*. Studies in Natural Language Processing. Cambridge, UK : Cambridge University Press.
- BAKER C. F., FILLMORE C. J. & LOWE J. B. (1998). The Berkeley FrameNet Project. In *Proceedings of the COLING-ACL*, Montreal, Canada.
- BARONI M., BERNARDINI S., FERRARESI A. & ZANCHETTA E. (2009). The wacky wide web : a collection of very large linguistically processed web-crawled corpora. *Language resources and evaluation*, **43**(3), 209–226.
- BRAUD C. (2011). Identification automatique des relations rhétoriques en français à partir de corpus annotés et de corpus bruts. Master's thesis, Université Paris Diderot.
- BRAUD C. & DENIS P. (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, p. 104–117, Sables d'Olonne, France.
- CANDITO M., CRABBÉ B. & DENIS P. (2010). Statistical french dependency parsing : Treebank conversion and first results. In *LREC*, Valletta, Malta.
- CARLETTA J. (1996). Assessing agreement on classification tasks : the kappa statistic. *Computational linguistics*, **22**(2), 249–254.
- CHAMBERS N. & JURAFSKY D. (2008). Unsupervised Learning of Narrative Event Chains. In *Proceedings of ACL-08 : HLT*, p. 789–797, Columbus, Ohio.
- CHKLOVSKI T. & PANTEL P. (2004). Verbocean : Mining the web for fine-grained semantic verb relations. In D. LIN & D. WU, Eds., *Proceedings of EMNLP 2004*, p. 33–40, Barcelona, Spain : Association for Computational Linguistics.
- DENIS P. & SAGOT B. (2012). Coupling an annotated corpus and a lexicon for state-of-the-art pos tagging. *Language Resources and Evaluation*, (46), 721–736.
- DO Q., CHAN Y. S. & ROTH D. (2011). Minimally supervised event causality identification. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, p. 294–303, Edinburgh, Scotland, UK. : Association for Computational Linguistics.
- EVERT S. (2005). *The statistics of word cooccurrences*. PhD thesis, Stuttgart University.
- FELBAUM C. (1998). *Wordnet, an Electronic Lexical Database for English*. Cambridge : MIT Press.
- FENG V. W. & HIRST G. (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)*, p. 60–68, Jeju Island, Korea : Association for Computational Linguistics.
- GREFENSTETTE G. (1994). *Explorations in automatic thesaurus discovery*. Springer.
- HASHIMOTO C., TORISAWA K., KURODA K., DE SAEGER S., MURATA M. & KAZAMA J. (2009). Large-scale verb entailment acquisition from the Web. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, p. 1172–1181, Singapore : Association for Computational Linguistics.
- HERNAULT H., PRENDINGER H., DUVERLE D. A. & ISHIZUKA M. (2010). HILDA : A Discourse Parser Using Support Vector Machine Classification. *Dialogue and Discourse*, **1**(3), 1–33.
- KOZAREVA Z. (2012). Cause-effect relation learning. In *Workshop Proceedings of TextGraphs-7 : Graph-based Methods for Natural Language Processing*, p. 39–43, Jeju, Republic of Korea : Association for Computational Linguistics.
- LIN D. (1998). Automatic retrieval and clustering of similar words. In *Proceedings of the 36th ACL and 17th COLING joint conference*, volume 2, p. 768–774, Montreal.
- LIN D. & PANTEL P. (2002). Concept discovery from text. In *Proceedings of Coling 2002*, p. 1–7 : Association for Computational Linguistics.
- LIN Z., NG H. T. & KAN M.-Y. (2014). A PDTB-styled end-to-end discourse parser. *Natural Language Engineering*, **20**(2), 151–184.

- LIU M., LI W., WU M. & LU Q. (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*, p. 185–188, Prague, Czech Republic : Association for Computational Linguistics.
- MANN W. C. & THOMPSON S. A. (1988). Rhetorical Structure Theory : Towards a Functional Theory of Text Organization. *Text*, **8**(3), 243–281.
- MARCU D. & ECHIHABI A. (2002). An Unsupervised Approach to Recognizing Discourse Relations. In *Proceedings of ACL*, p. 368–375.
- MIRROSHANDEL S. A., NASR A. & SAGOT B. (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*, p. 239–247, Atlanta, Georgia : Association for Computational Linguistics.
- NIVRE J., HALL J., NILSSON J., CHANEV A., ERYIGIT G., KÜBLER S., MARINOV S. & MARSI E. (2007). Malt-parser : A language-independent system for data-driven dependency parsing. *Natural Language Engineering*, **13**(2), 95–135.
- PRASAD R., DINESH N., LEE A., MILTSAKAKI E., ROBALDO L., JOSHI A. & WEBBER B. L. (2008). The Penn Discourse TreeBank 2.0. In *Proceedings of LREC 2008*.
- ROZE C., DANLOS L. & MULLER P. (2012). Lexconn : A french lexicon of discourse connectives. *Discours*, (10).
- SAGOT B. (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)*, Valletta, Malta.
- SPORLEDER C. & LASCARIDES A. (2008). Using Automatically Labelled Examples to Classify Rhetorical Relations : An Assessment. *Natural Language Engineering*, **14**(3), 369–416.
- TREMPER G. & FRANK A. (2013). A discriminative analysis of fine-grained semantic relations including presupposition : Annotation and classification. *Dialogue & Discourse*, **4**(2), 282–322.
- UZZAMAN N., LLORENS H., DERZYNSKI L., ALLEN J., VERHAGEN M. & PUSTEJOVSKY J. (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)*, p. 1–9, Atlanta, Georgia, USA : Association for Computational Linguistics.
- VAN DEN EYNDE K. & MERTENS P. (2010). Le dictionnaire de valence : Dicovalence. <http://bach.arts.kuleuven.be/dicovalence/>.
- WELLNER B., PUSTEJOVSKY J., HAVASI C., RUMSHISKY A. & SAURÍ R. (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*, p. 117–125, Stroudsburg, PA, USA : Association for Computational Linguistics.