

Université Paris Nanterre

Annotation automatique des adverbiaux temporels à l'aide d'Unitex

Par

CHENG Weixuan

HUANG Yidi

Enseignant : Delphine Battistelli
Master TAL Traitement automatique des langues
Unité d'enseignement : Modélisation linguistique pour
l'analyse automatique de textes
Année : 2022-2023

1. Introduction

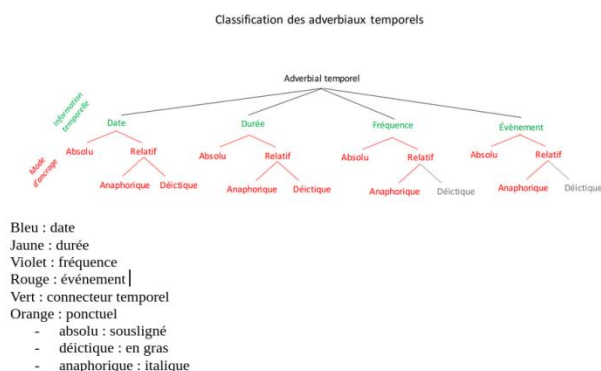
Les adverbes temporels sont une classe de modificateurs neutres qui reflètent le temps, la durée et la fréquence d'une action. En ce qui concerne les objets décrits par tous les adverbes temporels, d'une manière générale, si nous les divisons simplement, ils peuvent être grossièrement divisés en trois catégories : point de temps, période de temps et fréquence. Et ces trois catégories renvoient respectivement la date, la durée et la fréquence. Parmi eux, pour la grande catégorie de point temporel, en plus de la date la plus courante, il existe également la quatrième classe d'adverbes ponctuels qui sont précis à un certain nœud temporel mais ne sont pas exprimés à travers la date. De plus, il existe une classe (cinquième) un peu rare mais toujours existante, comme les adverbes temporels utilisés pour décrire et mettre en évidence des événements, soit historiques soit typiques. Ce type d'adverbe est généralement lié à un temps révélant un lieu temporel précis, appelé adverbe événementiel.

En plus d'exprimer simplement la temporalité, dans les énoncés, les adverbes de temps sont aussi une sorte de référence d'un point de vue pragmatique. Il contient trois modes de référence : « déictique » (indiquer une référence contextuelle), « anaphorique » (indiquer une référence « cotextuelle ») et « absolue » (indiquer une référence avec précision).

L'objectif de ce travail vise à repérer tous les adverbiaux temporels en fonction de leur type via l'Unitex et à comparer le résultat d'Unitex avec celui de l'annotation manuelle. Nous avons donc créé des dictionnaires des classes de mots et ensuite dessiné les graphes pour chaque type d'adverbe temporel.

2. Traitement du corpus

2.1 Annotations manuelles



(on a ajouté une catégorie : ponctuel)

Corpus 1 :

VILLEPIN PILE ET FACE ; Article paru dans l'édition du journal Le Monde du 4 Octobre 2005

1. **Le 21 juin dernier**, chaleur et gris sur Paris. A Matignon, des ouvriers montent une estrade pour la Fête de la musique. Dans son bureau-spa du 1er étage, Dominique de Villepin, en chemise et cravate bleue, long, mince, oppose un calme inédit à ceux qui tiennent son échec pour assuré : « Je voudrais prouver que le pouvoir n'est pas forcément le refuge du cynisme, du scepticisme et de l'inaction. » *Ce matin-là*, Thierry Breton, le ministre des finances, a parlé de la dette publique avec des chiffres abyssaux. *Trois semaines plus tôt*, la France a dit non à l'Europe, cette maison commune paradoxale, parce que non démocratique, dont les architectes semblent avoir *au fil du temps* oublié les origines et les fins. Une morosité durable affecte notre pays, dirigé *depuis longtemps* par des gouvernants paralysés par leur prudence. Dernier en titre : Jacques Chirac. Les Français, tentés de vivre en arrière, dans une quiétude provinciale, s'enferment avec leurs plaintes, s'adonnent à la peur et au repentir, sans savoir qu'ils sont enviés, leur volonté trébuche. « Notre démocratie est complètement bloquée, dit Villepin. Je ne peux réformer que par surprise, en restant dans l'équilibre, la vraie nature française, dans la justice, qui n'est pas l'égalité, et dans le mouvement. C'est seulement parce que c'est difficile que je peux réussir. »

Les deux plus proches collaborateurs de Villepin nous rejoignent. Il entretient avec eux un dialogue de chaque instant. Pierre Mongin, cheveux courts et gris, sourcils et yeux noirs, teint mat, bistre des cernes ; et Bruno Le Maire, grande taille, peau pâle, sourire et placidité sans affectation sur toute sa personne. Mongin a apporté une photo de Jean Moulin sur son bureau de directeur de cabinet. Sur celui de Le Maire (auteur d'une thèse sur Proust sous la direction du professeur Tadié), un exemplaire des *Essais* de Montaigne. Deux styles. Villepin reprend avec eux une discussion assez vive. « Vous avez les chiffres que Breton a donnés *ce matin*. Il faudra trouver les moyens de faire des économies. C'est juin 1940, nous sommes le dos au mur. Est-ce que les gens s'en rendent compte ? »

Depuis juillet 2004, les trois hommes se sont préparés à ce qu'ils considèrent comme une mission de la dernière chance. La Place Beauvau, *sous Villepin*, est devenue une sorte de laboratoire clandestin de la société française, où le ministre de l'intérieur a beaucoup reçu, écouté, sans jamais rien en laisser savoir. (...)

J'avais été surpris, Place Beauvau, de l'entendre dire qu'il faisait alors « un travail sur lui-même ». Curieux. Mais il avait le sentiment qu'il y avait quelque chose de fondamental dans notre société (les violences, les angoisses, les crispations, mais aussi l'identité nationale) qui lui échappait (et à tous les politiques) et qu'il ne pouvait imaginer réformer sans *d'abord* se réformer lui-même. (...)

2. **Le 15 septembre dernier**, le premier ministre français rit avec Zapatero, serre la main de Bush, embrasse Kofi Annan et Lula, parle avec Poutine, Jintao, Blair et Berlusconi. Il vient de passer une nuit blanche à mettre au point ses déclarations dans une chambre de l'hôtel Mandarin Oriental à Columbus Circle, avec Bruno Le Maire et nos deux ambassadeurs, Jean David Levitte et Jean-Marc de la Sablière. *Quand il arrive à la tribune*, George W. Bush se redresse sur son siège et branche son écouteur.

L'aisance avec laquelle Villepin a endossé les habits présidentiels pour s'asseoir à la table des grands et souffler avec eux les soixante bougies de l'ONU ne doit rien au hasard. Il y a vingt-cinq ans que Villepin est entré dans la diplomatie et qu'il en fréquente *chaque jour* les hommes et les dossiers.

A l'aube des années 1980, pour ce jeune homme né au Maroc *en 1953*, sortant de l'ENA, assoiffé de mouvement et qui n'oublie pas que de Gaulle a toujours raisonné, *dès juin 1940*, à l'échelle de la planète, la diplomatie était plus qu'une vocation, une évidence. D'autant que le Quai d'Orsay sait s'y prendre avec les poètes comme lui, ceux dont les mots roulent de la lave ou des délicatesses de sylphe. Il suffit que leurs dépêches restent concises. *Après Washington, New Delhi*. Dans l'air indien montent les fumées des bûches où des hommes s'immolent. Villepin est « ébloui » par la douceur d'un homme, Rajiv Gandhi, revenu affronter le chaos, *après une traversée du désert*. « Il m'a aidé à comprendre la grande leçon de Napoléon et de Gaulle : il y a toujours deux chances. » Puis c'est le retour à Paris *en 1992*, et à ses dossiers africains, au Quai. avoir un plus grand corpus afin de conclure autant que possible d'expressions.

C'est peu après que Chirac le convoque : « Dominique, lui dit-il, voyez Balladur. Il faut l'aider à préparer sa réflexion diplomatique. » Il se met *aussitôt* au travail avec Nicolas Bazire et Edouard Balladur, qui lui propose, *en arrivant à Matignon*, de devenir son conseiller diplomatique. *Trop tard*, Alain Juppé vient de lui demander de prendre la direction de son cabinet au Quai d'Orsay.

A la veille de Noël 1994, l'affaire de l'Airbus piraté à Alger le mobilise *vingt-quatre heures sur vingt-quatre*. Chirac, qui parle à Villepin au téléphone *dix fois par jour*, comprend que son interlocuteur est seul à Paris. « Qu'est-ce que vous faites *ce soir* ? », demande le maire de Paris. « Rien, je reste à mon bureau. » « Venez dîner avec nous. » « Et c'est ainsi, raconte Villepin, que je me suis retrouvé pour *le réveillon de Noël* chez Joe Allen, aux Halles, avec Chirac, Claude et Bernadette. » *Depuis le début de la prise d'otages*, Villepin plaide pour que l'avion soit autorisé à décoller et à atterrir en France, où la police trouvera plus facilement le moyen d'intervenir. *Le lendemain*, Chirac persuade Pasqua de laisser décoller l'avion pour Marseille.

En 2002, après cinq années au palais de l'Élysée, le voici de retour comme ministre des affaires étrangères. Contre toute attente, car il s'était préparé à partir pour la Place Beauvau. Sur sa table de travail, qui fut celle de Vergennes, l'encrier de Talleyrand, ce « diable boiteux » qui a *toujours* voulu être l'homme de la France, et près de lui, face à ses visiteurs, un tableau de Zao Wou-ki. Personne, et lui non plus, ne peut encore imaginer les prochaines accélérations de l'Histoire, qui vont lui donner un rôle.

L'Amérique, traumatisée par le 11-Septembre, cherche à entraîner la communauté internationale dans une embuscade diplomatique à l'ONU contre l'Irak, pays désarmé dont elle prétend qu'il menace la paix mondiale. *Le 14 février 2003*, Villepin porte la parole française à l'ONU. Son discours est applaudi dans l'enceinte des Nations unies. En France, il provoque un frisson à droite comme à gauche. Le pouvoir est aussi une question d'incarnation. Villepin a incarné, *ce jour-là*, une certaine idée de la France. En Amérique du Sud, où les télévisions diffusaient alors *quotidiennement* ses interviews en espagnol, dans les pays arabes et musulmans, en Afrique, il est devenu l'homme qui résiste à George W. Bush. Ce sont les événements qui fabriquent les hommes. Villepin est en phase avec l'idée qu'il s'est toujours faite de la vocation universelle de notre pays : il fonce.

Je l'ai alors accompagné *pendant deux mois*. Il menait les conseillers qui lui faisaient cortège, leurs impedimenta informatiques dans les bras, à un train de marathonien, les exhortant d'une voix forte à parfaire jusqu'à la dernière minute chacune des interventions qu'ils avaient souvent passé la nuit à préparer. Son principal ennemi ? (...)

Corpus 2

L'IMMEUBLE D'INES ET D'ADAM

Dans l'appartement d'Inès, 4 ans, et Adam, 2 ans, il y a des dizaines de photos de famille sur les murs. Des coussins moelleux sont disposés sur le canapé, et les enfants regardent des dessins animés à la télé. Rien d'anormal ! *Enfin* presque ... Cet appartement situé à Coulommiers, une ville à l'est de Paris, n'est pas vraiment le leur. Ils y ont emménagé en catastrophe avec leurs parents et leurs deux chats, *la veille de Noël*. L'immeuble où ils vivaient juste avant menaçait de s'effondrer... La famille était dans une situation de mal-logement.

« *Il y a deux ans et demi*, des fissures ont commencé à apparaître sur les murs de notre immeuble. Petit à petit, elles ont grossi. *A la fin*, on pouvait passer nos bras à l'intérieur ! », raconte Jennifer, la maman d'Inès et Adam. A cause de ces trous, il faisait *toujours* (déictique et anaphorique) très froid et humide chez eux. Adam tombait *souvent* (déictique et anaphorique) malade : « On l'emmenait *trois fois par mois* chez le pédiatre », affirme Jalil, son papa.

Inès aimait beaucoup sa chambre « violette et blanche avec des rideaux roses ». Mais elle se souvient des murs qui n'arrêtaient pas de faire « crac ». Ça lui faisait très peur. « La maison allait tomber sur nous ... », murmure-t-elle en se blottissant sur le canapé. Pour protéger leurs enfants, Jennifer et Jalil ont fini par les faire dormir chez leur grand-mère.

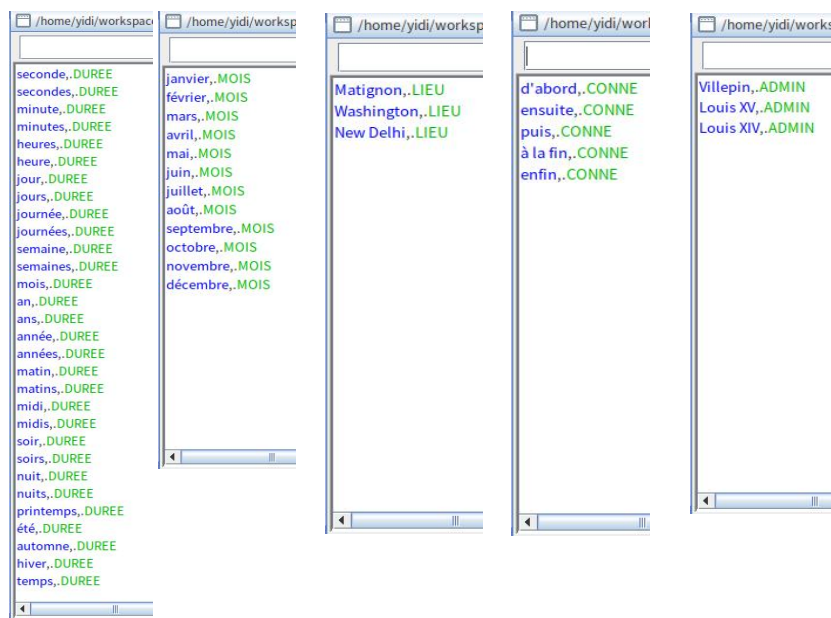
Les habitants de l'immeuble ont demandé des dizaines, voire peut-être des centaines de fois, aux propriétaires de faire des travaux, sans succès. *Cet automne*, Jennifer a appelé l'association Droit au logement, qui aide les personnes dans ce genre de situation. Une bénévole a alors tout mis en œuvre pour les sortir de là.

La maire de Coulommiers leur a trouvé un appartement de secours, dans lequel ils peuvent vivre gratuitement en attendant de trouver une solution. Même si leur combat pour trouver un nouveau logement est loin d'être terminé, toute la famille est déjà un peu soulagée. Inès la première : **“Maintenant, la nuit**, on a bien chaud”, se réjouit-elle.

2.2 Travail de préparation : création des dictionnaires

Le dictionnaire regroupe les syntagmes qui servent de la même fonction dans une structure temporelle, ce qui favorise le dessin de graphe en économisant le nombre de nœuds. Par exemple, le premier dictionnaire est composé des syntagmes qui implique une durée, qui construit une partie essentielle dans la plupart des adverbiaux temporels. En même temps, certains mots dépourvus de codes sémantiques et syntaxiques sont déjà rassemblés et attachés d'une étiquette, comme les pronoms démonstratifs « ce, cette, ... » possèdent une étiquette <Ddem>. On n'a plus besoin de créer un nouveau dictionnaire mais simplement de suivre la nomination du système Unitex.

En plus, il est à noter que, pour une catégorie spéciale des adverbiaux spatiaux : connecteur temporel, nous avons également créé un dictionnaire pour des recherches potentielles.



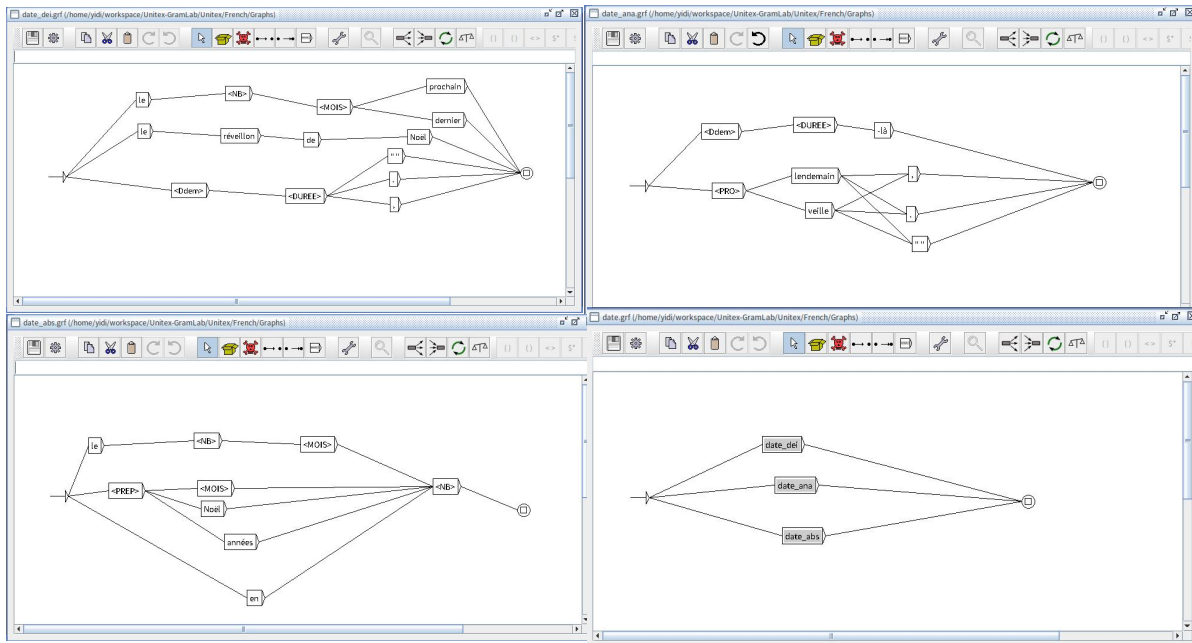
2.3 Graphes et sous-graphes catégorisées

Les graphes sont dessinés en fonction des adverbiaux temporels annotés manuellement. Dans chaque sous-catégorie, nous avons d'abord étudié les cas issus du texte 1. Lorsque les règles générales sont trouvées pour ces expressions, nous avons ensuite essayé de les visualiser à l'aide des nœuds et des arcs. Ainsi sont construits les sous-graphes et les graphes finales.

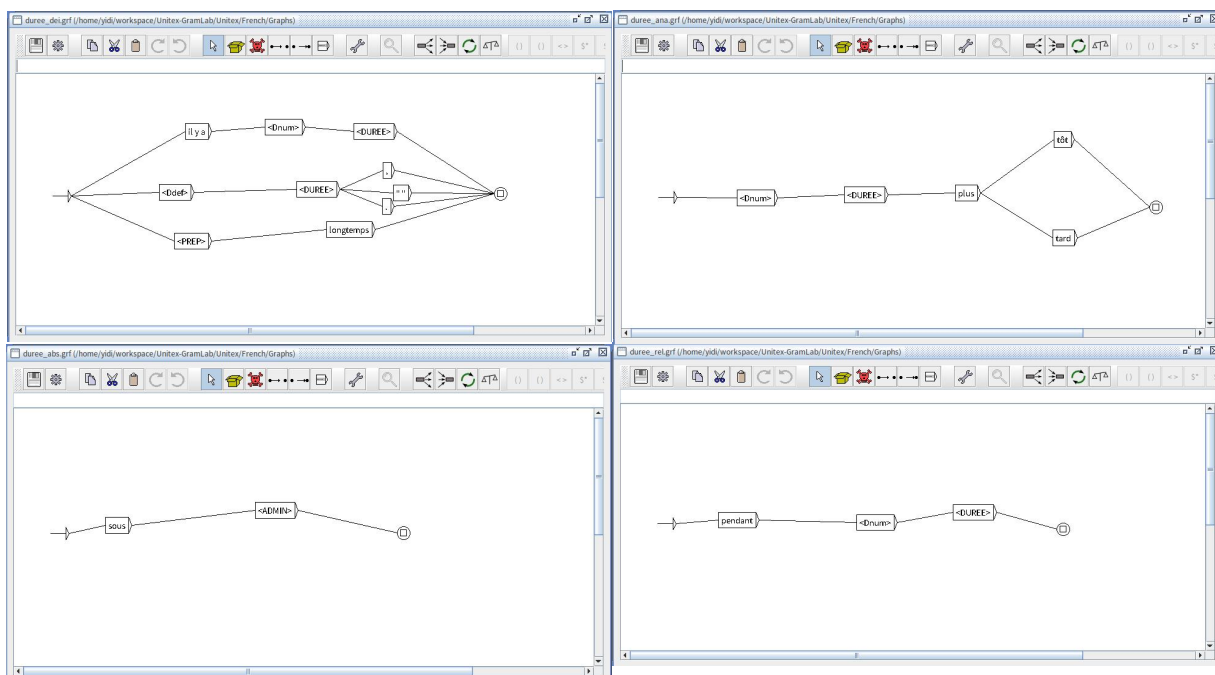
2.3.1 Date : déictique, anaphorique, absolu, total

Les trois nœuds parallèles : espace (“ ”), virgule (,) et point (.), servent à couper les syntagmes et distinguer les adverbiaux comme la date déictique « ce soir » de la date anaphorique « ce soir-là ».

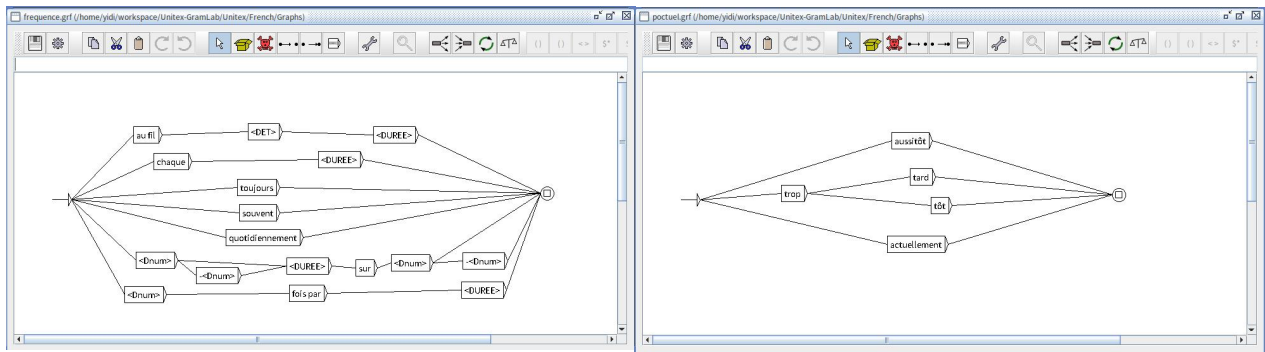
En plus, avec le format du nœud : « : » + nom de sous-graphe, il est possible de chercher les adverbiaux temporels de la même catégorie avec une graphe synthétique.



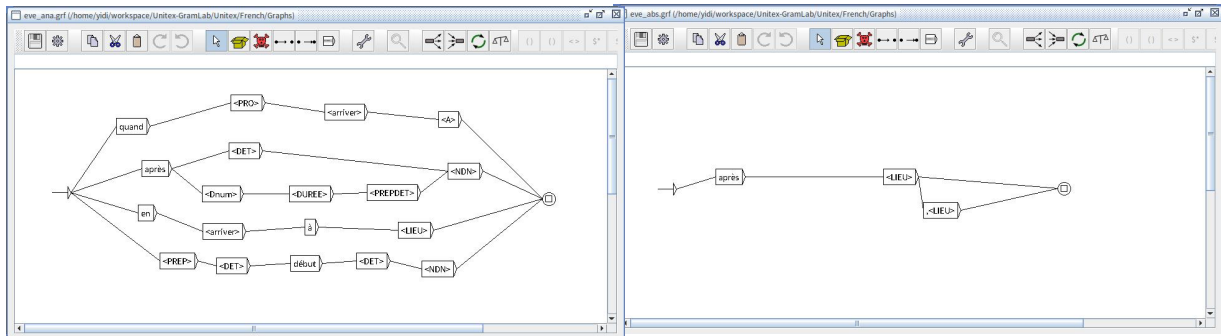
2.3.2 Durée : déictique, anaphorique, absolue, relatif (la graphe synthétique ne sera plus affichée)



2.3.3 Fréquence, Ponctuel



2.3.4 Événement : anaphorique, absolu



3. Evaluation

1) Résultats du 1er texte

<p>Concordance: /home/yidi/workspace/Unitex-GramLab/Unitex/French/Corpus/texteVillevin_snt/concord.html</p> <p>2 matches</p> <p>sa réflexion diplomatique. » Il se met aussitôt au travail avec Nicolas Bazire et Edouard Ball devenir son conseiller diplomatique.(S) Trop tard.(S) Alain Juppé vient de lui demander de pren</p>	<p>Concordance: /home/yidi/workspace/Unitex-GramLab/Unitex/French/Corpus/texteVillevin_snt/concord.html</p> <p>6 matches</p> <p>ler l'avion pour Marseille. (S)En 2002, après cinq années au palais de l'Elysée, le voici de re ajiv Gandhi, revenu affronter le chaos, après une traversée du désert. « Il m'a aidé à comprendre que leurs dépêches restent concises. (S)Après Washington, New Delhi.(S) Dans l'air indien monte s, avec Chirac, Claude et Bernadette. » Depuis le début de la prise d'otages, Villepin plaide p e et Edouard Balladur, qui lui propose, en arrivant à Matignon, de devenir son conseiller diplo Levitte et Jean-Marc de la Sablière.(S) Quand il arrive à la tribune, George W. Bush se redress</p>
<p>Concordance: /home/yidi/workspace/Unitex-GramLab/Unitex/French/Corpus/texteVillevin_snt/concord.html</p> <p>5 matches</p> <p>sité durable affecte notre pays, dirigé depuis longtemps par des gouvernants paralysés par leur erventions qu'ils avaient souvent passé la nuit à préparer.(S) Son principal ennemi ? (...) : il fonce. (S)Je l'ai alors accompagné pendant deux mois.(S) Il menait les conseillers qui lui a dernière chance.(S) La Place Beauvau, sous Villepin, est devenue une sorte de laboratoire cla publique avec des chiffres abyssaux.(S) Trois semaines plus tôt, la France a dit non à l'Europe</p>	<p>Concordance: /home/yidi/workspace/Unitex-GramLab/Unitex/French/Corpus/texteVillevin_snt/concord.html</p> <p>16 matches</p> <p>que jour les hommes et les dossiers. (S)A l'aube des années 1980, pour ce jeune homme né au M tion de son cabinet au Quai d'Orsay. (S)A la veille de Noël 1994, l'affaire de l'Airbus piraté n d'incarnation.(S) Villepin a incarné, ce jour-là, une certaine idée de la France.(S) En Améri sme, du scepticisme et de l'inaction. » Ce matin-là, Thierry Breton, le ministre des finances, s avez les chiffres que Breton a donnés ce matin-là.(S) Il faudra trouver les moyens de faire des ul à Paris. « Qu'est-ce que vous faites ce soir ? », demande le maire de Paris. « Rien, je rest e les gens s'en rendent compte ? » (S)Depuis juillet 2004, les trois hommes se sont préparés pas que de Gaulle a toujours raisonné, dès juin 1940, à l'échelle de la planète, la diplomatie s 1980, pour ce jeune homme né au Maroc en 1953, sortant de l'ENA, assoiffé de mouvement et q chances. » Puis c'est le retour à Paris en 1992, et à ses dossiers africains, au Quai. (S)C'est ser décoller l'avion pour Marseille. (S)En 2002, après cinq années au palais de l'Elysée, le vo étend qu'il menace la paix mondiale.(S) Le 14 février 2003, Villepin porte la parole française bord se réformer lui-même. (...) (S)2.(S) Le 15 septembre dernier, le premier ministre français r nal Le Monde du 4 Octobre 2005 (S)1.(S) Le 21 juin dernier, chaleur et gris sur Paris.(S) A Mat us facilement le moyen d'intervenir.(S) Le lendemain, Chirac persuade Pasqua de laisser décoller Villepin, que je me suis retrouvé pour le réveillon de Noël chez Joe Allen, aux Halles, avec C</p>

2) Résultats du 2ème texte

<p>Concordance: /home/yidi/workspace/Unitex-GramLab/Unitex/French/Corpus/Le Petit Libé_L' IMMEUBLE D' INE...</p> <p>2 matches</p> <p>it dans une situation de mal-logement. "Il y a deux ans et demi, des fissures ont commencé à ap gée.(S) Inès la première : "Maintenant, la nuit, on a bien chaud", se réjouit-elle.</p>	<p>Concordance: /home/yidi/workspace/Unitex-GramLab/Unitex/French/Corpus/Le Petit Libé_L' IMMEUBLE D' INE...</p> <p>2 matches</p> <p>s de faire des travaux, sans succès.(S) Cet automne, Jennifer a appelé l' association Droit au l avec leurs parents et leurs deux chats, la veille de Noël.(S) L' immeuble où ils vivaient juste</p>
<p>Concordance: /home/yidi/workspace/Unitex-GramLab/Unitex/French/Corpus/Le Petit Libé_L' IMMEUBLE D' INE...</p> <p>3 matches</p> <p>oid et humide chez eux.(S) Adam tombait souvent malade : "On l' emmenait trois fois par mois am.(S) A cause de ces trous, il faisait toujours très froid et humide chez eux.(S) Adam tombait tombait souvent malade : "On l' emmenait trois fois par mois chez le pédiatre", affirme Jaill,</p>	

3) Scores

	Texte 1	Texte 2
rappel	1	0,625
précision	1	0,714

La baisse de scores pour le texte 2 s'explique par le manque de ressources des adverbiaux temporels. Par exemple, la date déictique « maintenant » du texte 2 n'est pas enregistrée dans l'automate, ce qui entraîne la baisse du score de rappel.

Les expressions de durée comme « et demi » conduisent également à la réduction du score de précision. En résumé, il faut avoir un plus grand corpus afin de conclure autant que possible d'expressions des adverbiaux temporels et d'améliorer la qualité de l'automate.

4. Comparaison avec HeidelTime

Le résultat de HeidelTime montre que sa fonction de recherche se base sur les mots clé des adverbiaux temporels. Plus précisément, il cherche les syntagmes les plus courts. Par exemple, seulement la partie « ce soir » dans l'expression « ce soir-là » est marquée, ce qui produirait des problèmes lors de l'analyse potentielle de sous-catégorie : date déictique ou anaphorique.

Ce type de problèmes apparaît dans plusieurs catégories, mais arrive plus souvent pour les subordonnées temporelles à cause de la longueur de l'expression. Au contraire, notre analyse s'appuie sur les adverbiaux temporels parus dans le texte, les subordonnées complètes sont donc toutes annotées lors de la recherche.

En plus de cette différence essentielle, certains problèmes résident dans des catégories spécifiques. HeidelTime est assez performant pour trouver les expressions de date sous forme de « le + jour + mois + année » et les variants. La plupart de son résultat correspond bien à notre annotation manuelle et automatique. Cependant, l'automate rencontre la confusion entre la date et l'héméronyme (ex. juin 1940, 11-Septembre) : les héméronymes sont également annotés comme date. Ces deux derniers concepts demandent une analyse profonde sémantique, et c'est ce que manque l'automate. Alors la demande stricte de forme « le + ... » dans notre propre automate détourne ce problème.

Finalement, il est évident que le nombre d'annotations sous HeidelTime est inférieur à nos annotations. Cela s'explique probablement de son insensibilité aux adverbiaux temporels moins

évidents comme « Après Washington », « sous Villepin ». Il est possible que l'automate n'admette pas ces expressions comme adverbiaux temporels, ou il manque de telles règles pour l'annotation.

5. Conclusion

La comparaison des annotations manuelle et automatique montre que l'automate est assez capable de trouver les adverbiaux temporels sous forme fixe comme les dates. Cette application peut être perfectionnée avec l'enrichissement du corpus. L'exemple typique est que, l'expression de la date du texte 1 n'engage jamais les jours de semaine (lundi, mardi, etc.), le manque de ces ressources n'entraîne pas la baisse de scores dans texte 2, mais l'augmentation de ces adverbiaux complétera l'étendu de l'automate et réduira le biais de l'analyse des autres textes à l'avenir.

Les adverbiaux temporels sous forme typique sont relativement faciles à trouver, alors que certains adverbiaux moins évidents ou nécessitant une analyse sémantique constituent un obstacle pour la compréhension de la machine. Pour cette partie qui demande une étude profonde du contexte, l'automate est obligé à progresser afin de connaître des subordonnées (groupes prépositionnels, nominaux, etc.) dans les adverbiaux temporels et ainsi d'éviter les biais comme ceux de l'automate HeidelTime.