



Apports des analyses syntaxiques pour la détection automatique de mentions dans un corpus de français oral

Loïc Grobol, Isabelle Tellier, Éric Villemonte de la Clergerie, Marco Dinarelli, Frédéric Landragin

► To cite this version:

Loïc Grobol, Isabelle Tellier, Éric Villemonte de la Clergerie, Marco Dinarelli, Frédéric Landragin. Apports des analyses syntaxiques pour la détection automatique de mentions dans un corpus de français oral. TALN 2017, Association pour le Traitement Automatique des Langues (ATALA), Jun 2017, Orléans, France. hal-01558711

HAL Id: hal-01558711

<https://hal.inria.fr/hal-01558711>

Submitted on 9 Jul 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Apports des analyses syntaxiques pour la détection automatique de mentions dans un corpus de français oral

Loïc Grobol¹, Isabelle Tellier¹, Éric de la Clergerie²,
Marco Dinarelli¹, Frédéric Landragin¹

(1) Lattice (UMR 8094), CNRS, ENS Paris, Université Sorbonne Nouvelle,
PSL Research University, USPC, 1 rue Maurice Arnoux, 92120 Montrouge, France

(2) ALMAAnaCH, INRIA, 2 rue Simone Iff, 75589 Paris, France

loic.grobol@gmail.com, isabelle.tellier@sorbonne-nouvelle.fr,
eric.de_la_clergerie@inria.fr, marco.dinarelli@ens.fr, frederic.landragin@ens.fr

RÉSUMÉ

Cet article présente trois expériences de détection de mentions dans un corpus de français oral : ANCOR. Ces expériences utilisent des outils préexistants d'analyse syntaxique du français et des méthodes issues de travaux sur la coréférence, les anaphores et la détection d'entités nommées. Bien que ces outils ne soient pas optimisés pour le traitement de l'oral, la qualité de la détection des mentions que nous obtenons est comparable à l'état de l'art des systèmes conçus pour l'écrit dans d'autres langues. Nous concluons en proposant des perspectives pour l'amélioration des résultats que nous obtenons et la construction d'un système *end-to-end* pour lequel nos expériences peuvent servir de base de travail.

ABSTRACT

Experiences in using deep and shallow parsing to detect entity mentions in oral French

We present three experiments in detecting entity mentions in the corpus of oral French ANCOR, using publicly available parsing tools and state-of-the-art mention detection techniques used in coreference detection, anaphora resolution and Entity Detection and Tracking systems. While the tools we use are not specifically designed to deal with oral French, our results are comparable to those of state-of-the-art end-to-end systems for other languages. We also mention several ways to improve our results for future work in developing an end-to-end coreference resolution system for French, to which these experiments could be a baseline for mention detection.

MOTS-CLÉS : corpus de dialogues, coréférence, détection de mentions, analyse syntaxique, chunks, apprentissage automatique.

KEYWORDS : Dialogue corpus, Coreference, Mention detection, Parsing, Chunking, Machine Learning.

1 Introduction

La reconnaissance automatique des chaînes de coréférence dans un texte est une tâche complexe et fondamentale pour avoir accès au contenu informationnel d'un long document. Elle a donné lieu à de grandes campagnes d'évaluation, comme celles proposées par MUC (*Message Understanding Conference*), ACE (*Automatic Content Extraction*), SemEval (*Semantic Evaluation*) ou CoNLL (*Computational Natural Language Learning*) et à de nombreux travaux. Cette tâche est souvent présentée comme revenant à partitionner l'ensemble des mentions d'un texte. Mais, avant de regrouper entre elles les mentions coréférentes, il faut les identifier. C'est sur cette étape préliminaire que nous nous concentrons dans cet article.

Commençons par décrire plus précisément l'objet de notre étude : on appelle *mention* toute expression

faisant référence à une entité¹. Les mentions sont généralement des groupes nominaux ou des pronoms, mais tous les groupes nominaux et tous les pronoms ne sont pas des mentions. Ainsi, *il* dans « *il* y a pas encore de texte » et *la bienvenue* dans « la question là est la bienvenue et » sont non-référentiels et ne sont donc pas considérés comme des mentions. Détecter des mentions ne saurait donc se réduire à identifier les groupes nominaux et les pronoms. Les mentions peuvent également être imbriquées les unes dans les autres, voir par exemple dans la figure 1 « [le langage parlé et courant [des gens]] ».

On appelle *chaîne de coréférences* l'ensemble des mentions faisant référence à une même entité. On admet qu'une chaîne de coréférences peut n'être constituée que d'une seule mention² et on dira dans ce cas qu'il s'agit d'un *singleton*. La figure 1 montre un exemple de chaîne de coréférences (en gras) et des exemples de mentions (entre crochets).

- euh [on] dit quelquefois que [**la langue française**] se dégrade qu'est-ce que [vous] en pensez ?
- [**la langue française**] se dégrade dans [le langage parlé et courant [des gens]] [**elle**] se dégrade ?

FIG. 1: Mentions et chaînes de coréférences (corpus ANCOR).

On appelle *système complet d'identification de chaînes de coréférences* ou *système end-to-end* tout système capable de repérer les mentions dans un texte brut et de les regrouper en chaînes de coréférences. De tels systèmes ont déjà été conçus pour le français par (Trouilleux, 2001) et (Longo, 2013) mais leur portée reste modeste par rapport à celles de systèmes comme BART (Versley *et al.*, 2008) ou dcoref (Lee *et al.*, 2013), faute notamment de corpus en français annoté en chaînes de coréférences.

Un système — CROC — réalisant le regroupement des mentions en chaînes a déjà été réalisé pour le français par (Désoyer *et al.*, 2015) avec des performances comparables à l'état de l'art en anglais. Cependant, CROC s'appuyait sur des mentions de référence déjà repérées et portant des annotations morphosyntaxiques et sémantiques. Pour compléter ce travail, il reste donc à détecter les mentions dans le texte brut — c'est l'objet des travaux que nous présentons — et à y ajouter les annotations utilisées par CROC.

La détection des mentions n'est pas une priorité dans l'état de l'art : peu de travaux lui sont spécifiquement consacrés, beaucoup de systèmes *end-to-end* se contentent d'heuristiques simples, et certains travaux d'identification de chaînes de coréférence n'ont été développés qu'en tenant compte de mentions de référence.

De fait, les résultats obtenus par des systèmes de détection de mentions relativement simples sont globalement satisfaisants, ou en tout cas de suffisamment bonne qualité compte tenu de l'efficacité des systèmes de détection de coréférences qu'ils alimentent. Ces derniers, comportant généralement des détecteurs d'expressions non-référentielles, sont par ailleurs peu influencés par la précision de la détection de mentions, l'objectif est donc simplement de maximiser le rappel. La phase de détection des mentions des systèmes existants se résume donc bien souvent à la détection des groupes nominaux et des pronoms, tâche somme toute relativement simple pour des applications au langage écrit.

Dans notre cas, il serait cependant imprudent de négliger cette tâche. Le corpus avec lequel nous travaillons est en effet un corpus d'oral transcrit, et il n'est pas si évident que les outils d'analyse existants obtiennent des performances suffisantes pour lui appliquer directement les méthodes classiques. En effet, des travaux comme (Bove, 2008) ou (Tellier *et al.*, 2014) montrent que des techniques conçues pour l'écrit peuvent voir leurs performances significativement dégradées dans des applications directes au traitement de l'oral. Notre objectif est donc ici de déterminer dans quelle mesure les techniques classiques de détection de mentions peuvent s'appliquer au traitement du français oral et de produire une base de travail pour des recherches futures sur cette question.

1. Au sens de (Recasens, 2010)

2. Ce n'est pas le parti pris notamment pour les jeux de données des tâches MUC, CoNLL-2011 et CoNLL-2012, mais les singletons étant en fait très majoritaires parmi les chaînes de coréférences, en tenir compte permet d'améliorer significativement la détection des autres chaînes de coréférences (de Marneffe *et al.*, 2015).

2 Détection des mentions

2.1 Méthodes de détection

Nous présentons ici trois méthodes de détection des mentions dans un texte non-annoté. Nous les avons étudiées séparément, mais elle ne sont en rien mutuellement exclusives — (dos Santos & Carvalho, 2011) combine par exemple une extraction depuis une analyse syntaxique avec un système de détection d’entités nommées. Nous ne prétendons pas ici à l’exhaustivité : on pourra se reporter à (Poesio *et al.*, 2016) pour plus de détails.

2.1.1 Extraction à partir d’une analyse syntaxique

Cette méthode est utilisée par la majorité des systèmes *end-to-end* existants, généralement complétée par des systèmes de détection d’entités nommées et d’expressions non-référentielles. Elle consiste à extraire d’une analyse syntaxique les constituants pouvant être des mentions. Cette classification peut être réalisée à l’aide de règles écrites à la main (dos Santos & Carvalho, 2011), (Kummerfeld *et al.*, 2011) ou apprise automatiquement (Uryupina & Moschitti, 2013).

Cette procédure est raisonnablement fiable (voir tableau 1) et particulièrement adaptée au traitement de corpus arborés, comme ceux issus d’OntoNotes. Elle est cependant plus difficile à mettre en œuvre dans le cadre d’un système *end-to-end*, qui devient de ce fait dépendant de l’existence d’analyseurs syntaxiques pour la langue traitée et doit composer avec leurs éventuelles erreurs d’analyse.

2.1.2 Extraction à partir d’une analyse de surface

Les inconvénients de l’utilisation d’analyses syntaxiques évoqués dans la section 2.1.1 peuvent être — dans une certaine mesure — évités en utilisant non plus une analyse syntaxique complète, mais une analyse syntaxique de surface de type *chunking* (Abney, 1992). C’est le parti pris par (Soraluze *et al.*, 2012), qui utilise des outils de *chunking* et de détection de propositions dont les résultats sont ensuite traités par des règles écrites manuellement. L’avantage de cette approche est la plus grande disponibilité de *chunkers* de qualité et leur plus grande robustesse. L’inconvénient majeur est que ce type d’analyse, qui ignore les imbrications de constituants, ne permet pas de récupérer directement tous les types de mentions, et nécessite un traitement a posteriori — par exemple des règles de regroupement de *chunks*.

2.1.3 Apprentissage automatique de la détection des frontières de mentions

Les méthodes évoquées précédemment peuvent donner de bons résultats, mais sont entièrement dépendantes de l’existence de ressources linguistiques adaptées. Il existe cependant une autre possibilité : détecter directement dans le texte brut les frontières de mentions, comme le ferait un *chunker* ou un système de *NER*.

Cette technique a été étudiée entre autres par (Jing *et al.*, 2003), (Florian *et al.*, 2010), et (Qian *et al.*, 2007) (Nguyen *et al.*, 2016) pour la tâche de reconnaissance et de suivi d’entités (EDT) issue des campagnes ACE, qui ne traite de la détection de chaînes de coréférences que pour une classe restreinte d’entités. Elle permet de ramener le problème de la détection de mentions à une tâche d’étiquetage de séquences, pour laquelle des méthodes efficaces sont connues et ont été étudiées en profondeur : MEMM pour (Florian *et al.*, 2010), CRF pour (Qian *et al.*, 2007) et plus récemment réseaux de neurones récurrents pour (Nguyen *et al.*, 2016). Elle a en revanche le défaut classique de la modélisation de frontières par des étiquettes : la prise en compte d’entités imbriquées complique significativement les modèles (voir par exemple (Finkel & Manning, 2009)).

2.2 Mesures

Contrairement aux mesures de qualité de la détection des chaînes de coréférences, la mesure de la qualité de la détection des mentions tend à faire consensus et est en général donnée par le triplet

précision/rappel/ F_1 -mesure. La priorité est souvent donnée (par exemple dans (Lassalle, 2015)) au rappel au détriment de la précision, considérant que les faux positifs seront de toute façon détectés pendant la phase de détection des chaînes de coréférences. Il reste toutefois à déterminer ce que signifie « être correctement identifié » pour une mention trouvée par un système. (Recasens & Hovy, 2011) suggère deux interprétations :

Correspondance stricte une mention système correspond à une mention de référence si elle recouvre exactement le même empan de texte.

Correspondance indulgente une mention système S correspond à une mention de référence G si

1. S est inclus dans G
2. La tête syntaxique de G appartient à S

3 Expérimentations et résultats

Nous présentons ici les différentes expériences de détection des mentions que nous avons menées. Le tableau 1 présente (dans sa seconde partie) les résultats de ces expériences, dont les modalités sont détaillées par la suite, et — a titre indicatif — ceux de systèmes de référence. Notons cependant que ces systèmes n’ayant pas été développés dans les mêmes conditions, ni même pour les mêmes langues, une comparaison strictement quantitative aurait peu de sens. De même, nos expériences utilisant un apprentissage automatique n’ont été évaluées que sur 10 % d’ANCOR (le reste ayant servi à l’apprentissage), leurs résultats sont donc difficilement comparables avec ceux de nos autres expériences.

TABLE 1: Comparaison de systèmes de détection de mentions (correspondances strictes)

| Système | Corpus | P(%) | R(%) | F_1 (%) | Méthode |
|------------------------------------|--------------------------------|-------------------|-------------------|-------------------|----------------------|
| (Lassalle, 2015) | CoNLL-2012 (en) ¹ | 43,77 | 97,97 | 60,05 | 2.1.1 ^{3,2} |
| (Kummerfeld <i>et al.</i> , 2011) | OntoNotes (en) ¹ | 56,97 | 69,77 | 62,72 | 2.1.1 ² |
| (Ogrodniczuk <i>et al.</i> , 2014) | NKJP | 66,04 | 63,99 | 65,00 | 2.1.2 |
| (Soraluze <i>et al.</i> , 2012) | EPEC | 76,85 | 78,59 | 77,58 | 2.1.2 |
| (Uryupina & Moschitti, 2013) | CoNLL-2012 (ar) ^{1,5} | 66,0 | 66,1 | 66,1 | 2.1.1 ^{4,2} |
| (Uryupina & Moschitti, 2013) | CoNLL-2012 (zh) ^{1,5} | 68,9 | 71,3 | 70,1 | 2.1.1 ^{4,2} |
| (Uryupina & Moschitti, 2013) | CoNLL-2012 (ar) ^{1,5} | 31,07 | 90,67 | 46,28 | 2.1.1 ^{3,2} |
| (Nguyen <i>et al.</i> , 2016) | ACE 2005 ⁵ | 83,7 ⁶ | 81,8 ⁶ | 82,7 ⁶ | 2.1.3 |
| Analyse syntaxique | ANCOR | 57,28 | 77,07 | 65,72 | 2.1.1 |
| Analyse syntaxique | ANCOR ⁵ | 48,66 | 65,46 | 55,82 | 2.1.1 |
| Chunks | ANCOR | 45,99 | 36,78 | 40,87 | 2.1.2 |
| Apprentissage | ANCOR (test) | 89,09 | 20,04 | 32,72 | 2.1.3 |
| Apprentissage | ANCOR (test) ⁵ | 89,09 | 88,61 | 88,85 | 2.1.3 |

1. Ne tient pas compte des singletons
2. Utilise une analyse syntaxique de référence
3. Optimisé pour le rappel
4. Optimisé pour la précision
5. Sans mentions imbriquées
6. Tient compte de la catégorisation des mentions

3.1 Corpus

Nos travaux ont été effectués sur le corpus ANCOR (Muzerelle *et al.*, 2013), premier et actuellement seul corpus français annoté en chaînes de coréférences. Ce corpus a pour particularité d’être construit à partir d’oral conversationnel transcrit, ce qui pose un défi supplémentaire pour l’utilisation des méthodes évoquées dans la section 2.1, toutes développées pour le traitement du langage écrit. Une autre particularité notable par rapport à une grande partie des corpus utilisés dans l’état de l’art est qu’il n’existe à l’heure actuelle aucune analyse syntaxique de référence pour ANCOR. De ce fait, il ne nous est pas possible d’évaluer les performances en correspondance indulgente des méthodes que nous avons expérimentées.

3.2 Extraction depuis des analyses syntaxiques

Nous utilisons dans cette approche les analyses syntaxiques fournies par Talismane (Urieli, 2013), un analyseur statistique en dépendances pour le français. La méthode que nous utilisons est la suivante :

1. Repérer dans les sorties de l’analyseur les mots des catégories suivantes : noms communs (nc), noms propres (npp), clitiques sujets (cls), clitiques objets (clo), pronoms personnels (pro), pronoms relatifs (prorel), adjectifs numériques (num). Ces mots seront les têtes des futurs candidats-mentions.
2. Pour chaque tête ainsi repérée, déterminer son extension maximale, en suivant récursivement les relations de dépendances.

Les résultats obtenus sont les suivants : $P \approx 57,28\%$, $R \approx 77,07\%$ et $F_1 \approx 65,72\%$.

Si la précision obtenue n’est pas entièrement satisfaisante, le rappel est comparable à l’état de l’art, même si la diversité des conventions d’annotation des mentions rend la comparaison difficile. Une analyse des erreurs produites par cette méthode indique qu’un mauvais traitement des disfluences est la plus grande cause de faux positifs, il paraît donc raisonnable de supposer que des heuristiques de rectification de frontières *a posteriori* amélioreraient significativement à la fois la précision et le rappel. De même, utiliser un analyseur syntaxique plus performant pour l’oral serait sans doute très profitable, nos résultats étant entièrement dépendants de la qualité de l’analyse syntaxique. La plupart des travaux cités dans le tableau 1 utilisent d’ailleurs des analyses syntaxiques manuelles, ou au moins corrigées manuellement plutôt que des analyses automatiques brutes. Cela dit, il reste parmi les candidats-mentions ainsi repérés un certain nombre d’expressions non-référentielles, notamment des *ils* impersonnels, qui rendent de toute façon nécessaire l’ajout en sortie d’un filtre adapté, utilisant par exemple ILIMP (Danlos, 2005).

3.3 Extraction depuis les chunks

Pour cette approche, nous utilisons les chunks nominaux (NP) détectés par SEM (Tellier *et al.*, 2012) directement comme candidats-mentions. Plus spécifiquement, nous utilisons le mode de détection des seuls NP, qui permet en réalité de détecter aussi des NP qui seraient inclus dans des chunks prépositionnels. Les modèles utilisés sont les modèles standards fournis avec SEM, appris sur le French Treebank (Abeillé *et al.*, 2003).

Les résultats obtenus par cette méthode sont les suivants : $P \approx 45,99\%$, $R \approx 36,78\%$, $F_1 \approx 40,87\%$.

Ces résultats, de prime abord assez décevants, sont toutefois à considérer en tenant compte des remarques faites dans la section 2.1.2 : les chunks ne tenant par définition pas compte des imbrications, ils ne peuvent représenter qu’une fraction des mentions et doivent nécessairement être traités et combinés pour reconstituer l’ensemble des mentions — et ce même dans l’hypothèse d’un chunking parfait. D’ailleurs, la qualité du chunking utilisé ici est loin d’être garantie, pour mémoire, (Tellier *et al.*, 2014) évaluait aux alentours de 77 % seulement la F_1 -mesure du chunking de obtenu par les modèles standards de SEM sur un échantillon du corpus oral ESLO, à partir duquel est construit la plus grande partie d’ANCOR. Il n’est donc pas aberrant d’imaginer que la qualité des chunks nominaux que nous utilisons ici est largement sous-optimal, et que — là encore — l’utilisation de modèles spécifiquement appris pour le traitement de l’oral comme ceux décrits dans (Tellier *et al.*, 2014) pourraient grandement améliorer ces performances.

Du fait de ces limitations, il ne semble donc pas pertinent pour l’instant de comparer ces résultats avec ceux de travaux similaires, comme par exemple (Soraluze *et al.*, 2012).

3.4 Détection directe des frontières

La dernière méthode que nous avons expérimentée a consisté à apprendre directement le repérage des frontières de mentions, en le traitant comme une tâche d’étiquetage de séquences. Spécifiquement, en considérant chaque tour de parole du corpus comme une séquence de tokens, nous avons entraîné un modèle pour l’étiqueteur à CRF linéaire Wapiti (Lavergne *et al.*, 2010) associant à chaque token une étiquette du schéma BILOU (*Begin, Inside, Last, Outside, Unique*) utilisé classiquement pour l’extraction d’information par annotation, par exemple par (Nguyen *et al.*, 2016) pour la reconnaissance de mentions. Comme nous l’avons précisé en section 2.1.3, cette méthode ne permet pas non plus de prendre en compte les mentions imbriquées. Nous avons donc fait le choix de ne conserver pour cette expérience que les mentions les plus larges et d’ignorer leurs sous-mentions.

La procédure d’apprentissage a été réalisée sur une partition arbitraire du corpus ANCOR, suivant la répartition classique de 80%/10%/10% respectivement pour les sous-corpus d’apprentissage (pour l’entraînement de l’étiqueteur), de développement (utilisé comme critère d’arrêt de l’apprentissage) et d’évaluation. En plus des formes des tokens, nous avons également utilisé comme traits pour l’apprentissage les parties du discours et les frontières de chunks et d’entités nommées déterminées par SEM.

Les résultats sont les suivants : $P \approx 89,09\%$, $R \approx 88,61\%$ et $F_1 \approx 88,85\%$ en ne tenant compte que des mentions maximales. On obtient $R \approx 20,04\%$ et $F_1 \approx 32,72\%$ en tenant compte de l’ensemble des mentions.

Ces résultats sont raisonnablement satisfaisants même si la comparaison avec l’état de l’art est délicate, les travaux en détection de coréférences prenant habituellement en compte toutes les mentions — y compris imbriquées — et les systèmes de suivi d’entités tenant en général compte de la catégorisation des entités en plus de leur repérage dans leurs évaluations.

4 Conclusion et perspectives

La détection automatique de chaînes de coréférences en français n’en est encore qu’à ses débuts, la mise à disposition de corpus annotés étant relativement récente. Le défi que représente le développement d’un système *end-to-end* adapté à cette nouvelle langue et au contexte spécifique de l’oral impose l’étude approfondie de cette tâche parfois négligée. Les expériences que nous avons exposées ici fournissent une base de travail pour des travaux futurs et montrent que des méthodes classiques exploitant des analyses syntaxiques peuvent donner des résultats encourageants, même dans ce nouveau contexte. Il reste à présent à bâtir sur ces fondations un système de détection des mentions capable de s’interfacer avec un système de détection de coréférences pour disposer d’un système *end-to-end* pour le français — par exemple en le nourrissant de traits issus des analyses syntaxiques utilisées pour la détection. La précision de la détection des mentions devra également être améliorée — et ce quelle que soit la méthode employée — en filtrant *a posteriori* les expressions non-référentielles erronément détectées.

Dans cette perspective, des travaux futurs devront probablement exploiter des ressources linguistiques plus adaptées — notamment au français oral — et employer des méthodes plus puissantes que celles que nous avons expérimentées ici, qu’il s’agisse de règles écrites à la main ou d’apprentissage automatique. Il serait ainsi intéressant de combiner plusieurs de ces approches dans un système hybride, ou de les appliquer en cascade — par exemple en appliquant des analyseurs syntaxiques directement sur les mentions maximales détectées par apprentissage. On pourrait également envisager de traiter conjointement détection des mention et détection des coréférences.

5 Remerciements

Cette recherche s'insère dans le programme « Investissements d'Avenir » géré par l'Agence Nationale de la Recherche ANR-10-LABX-0083 (Labex EFL).

Ce travail a par ailleurs bénéficié du soutien de l'ANR DEMOCRAT (Description et modélisation des chaînes de référence : outils pour l'annotation de corpus et le traitement automatique), projet ANR-15-CE38-0008.

Références

- ABEILLÉ A., CLÉMENT L. & TOUSSENEL F. (2003). *Building a Treebank for French*, In A. ABEILLÉ, Ed., *Treebanks : Building and Using Parsed Corpora*, p. 165–187. Springer Netherlands : Dordrecht.
- ABNEY S. P. (1992). *Parsing By Chunks*, In R. C. BERWICK, S. P. ABNEY & C. TENNY, Eds., *Principle-Based Parsing : Computation and Psycholinguistics*, p. 257–278. Springer Netherlands : Dordrecht.
- BOVE R. (2008). *Automatic speech parsing : study of disfluencies*. PhD thesis, Université de Provence - Aix-Marseille I.
- DANLOS L. (2005). ILIMP : Outil pour repérer les occurrences du pronom impersonnel il. In *Actes de la 12ème conférence sur le Traitement Automatique des Langues Naturelles*, p. 123–132, Dourdan, France : Association pour le Traitement Automatique des Langues.
- DE MARNEFFE M.-C., RECASENS M. & POTTS C. (2015). Modeling the lifespan of discourse entities with application to coreference resolution. *Journal of Artificial Intelligence Research*, **52**, 445–475.
- DÉSOYER A., LANDRAGIN F., TELLIER I., LEFEUVRE A. & ANTOINE J.-Y. (2015). Coreference Resolution for Oral Corpus : a machine learning experiment with ANCOR corpus. *Traitement Automatique des Langues*, **55**(2), 97–121.
- DOS SANTOS C. N. & CARVALHO D. L. (2011). Rule and tree ensembles for unrestricted coreference resolution. In *Proceedings of the Fifteenth Conference on Computational Natural Language Learning : Shared Task*, CONLL Shared Task '11, p. 51–55, Stroudsburg, PA, USA : Association for Computational Linguistics.
- FINKEL J. R. & MANNING C. D. (2009). Nested named entity recognition. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing : Volume 1 - Volume 1*, EMNLP '09, p. 141–150, Stroudsburg, PA, USA : Association for Computational Linguistics.
- FLORIAN R., PITRELLI J. F., ROUKOS S. & ZITOUNI I. (2010). Improving mention detection robustness to noisy input. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, EMNLP '10, p. 335–345, Stroudsburg, PA, USA : Association for Computational Linguistics.
- JING H., FLORIAN R., LUO X., ZHANG T. & ITTYCHERIAH A. (2003). HowtogetaChineseName(Entity) : Segmentation and combination issues. In *Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing*, EMNLP '03, p. 200–207, Stroudsburg, PA, USA : Association for Computational Linguistics.
- KUMMERFELD J. K., BANSAL M., BURKETT D. & KLEIN D. (2011). Mention detection : Heuristics for the ontonotes annotations. In *Proceedings of the Fifteenth Conference on Computational Natural Language Learning : Shared Task*, CONLL Shared Task '11, p. 102–106, Stroudsburg, PA, USA : Association for Computational Linguistics.
- LASSALLE E. (2015). *Structured learning with latent trees : a joint approach to coreference resolution*. PhD thesis.
- LAVERGNE T., CAPPÉ O. & YVON F. (2010). Practical very large scale CRFs. In *Proceedings the 48th Annual Meeting of the Association for Computational Linguistics (ACL)*, p. 504–513 : Association for Computational Linguistics.
- LEE H., CHANG A., PEIRSMAN Y., CHAMBERS N., SURDEANU M. & JURAFSKY D. (2013). Deterministic coreference resolution based on entity-centric, precision-ranked rules. *Computational Linguistics*, **39**(4), 885–916.

- LONGO L. (2013). *Vers des moteurs de recherche "intelligents" : un outil de détection automatique de thèmes. Méthode basée sur l'identification automatique des chaînes de référence*. PhD thesis, Université de Strasbourg.
- MUZERELLE J., LEFEUVRE A., ANTOINE J.-Y., SCHANG E., MAUREL D., VILLANEAU J. & ESHKOL I. (2013). Ancor, premier corpus de français parlé d'envergure annoté en coréférence et distribué librement. In *Actes de la 20e conférence sur le Traitement Automatique des Langues Naturelles*, p. 555–563, Les Sables d'Olonne, France : Association pour le Traitement Automatique des Langues.
- NGUYEN T. H., SIL A., DINU G. & FLORIAN R. (2016). Toward mention detection robustness with recurrent neural networks. *CoRR*, **abs/1602.07749**.
- OGRODNICZUK M., WÓJCICKA A., GŁOWIŃSKA K. & KOPEĆ M. (2014). *Detection of Nested Mentions for Coreference Resolution in Polish*, In A. PRZEPIÓRKOWSKI & M. OGRODNICZUK, Eds., *Advances in Natural Language Processing : 9th International Conference on NLP, PolTAL 2014, Warsaw, Poland, September 17-19, 2014. Proceedings*, p. 270–277. Springer International Publishing : Cham.
- POESIO M., STUCKARDT R. & VERSLEY Y. (2016). *Anaphora Resolution : Algorithms, Resources, and Applications*. Theory and Applications of Natural Language Processing. Springer Berlin Heidelberg.
- QIAN D., LI W., YUAN C., LU Q. & WU M. (2007). Applying machine learning to Chinese entity detection and tracking. In *Proceedings of the 8th International Conference on Computational Linguistics and Intelligent Text Processing, CICLing '07*, p. 154–165, Berlin, Heidelberg : Springer-Verlag.
- RECASENS M. (2010). *Coreference : Theory, Annotation, Resolution and Evaluation*. PhD thesis.
- RECASENS M. & HOVY E. (2011). Blanc : Implementing the Rand index for coreference evaluation. *Nat. Lang. Eng.*, **17**(4), 485–510.
- SORALUZE A., ARREGI O., ARREGI X., CEBERIO K. & DE ILARRAZA A. D. (2012). Mention detection : First steps in the development of a Basque coreference resolution system. In J. JANCSARY, Ed., *Proceedings of KONVENS 2012*, p. 128–136 : ÖGAI.
- TELLIER I., DUCHIER D., ESHKOL I., COURMET A. & MARTINET M. (2012). Apprentissage automatique d'un chunker pour le français. In *Actes de la 19e conférence sur le Traitement Automatique des Langues Naturelles*, p. 431–438, Grenoble, France : Association pour le Traitement Automatique des Langues.
- TELLIER I., ESHKOL-TARAVELLA I., DUPONT Y. & WANG I. (2014). Peut-on bien chunker avec de mauvaises étiquettes POS ? In *Actes de la 21e conférence sur le Traitement Automatique des Langues Naturelles*, p. 125–136, Marseille, France : Association pour le Traitement Automatique des Langues.
- TROUILLEUX F. (2001). *Referential links identification and automatic interpretation of pronominal expressions in French texts*. PhD thesis, Université Blaise-Pascal, Clermont-Ferrand.
- URIELI A. (2013). *Robust French syntax analysis : reconciling statistical methods and linguistic knowledge in the Talismane toolkit*. PhD thesis, Université de Toulouse II le Mirail.
- URYUPINA O. & MOSCHITTI A. (2013). Multilingual mention detection for coreference resolution. In *IJCNLP*, p. 100–108.
- VERSLEY Y., PONZETTO S. P., POESIO M., EIDELMAN V., JERN A., SMITH J., YANG X. & MOSCHITTI A. (2008). BART : A modular toolkit for coreference resolution. In *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics on Human Language Technologies : Demo Session, HLT-Demonstrations '08*, p. 9–12, Stroudsburg, PA, USA : Association for Computational Linguistics.