

Résolution des coréférences neuronale : une approche basée sur les têtes

Quentin Gliosca^{1,2*} Pascal Amsili³

¹ École Polytechnique Fédérale de Lausanne

² Haute École d'Ingénierie et de Gestion du Canton de Vaud

³ Laboratoire de Linguistique Formelle (Université Paris Diderot & CNRS)

¹ quentin.gliosca@gmail.com

³ amsili@linguist.univ-paris-diderot.fr

RÉSUMÉ

L'avènement des approches neuronales de bout en bout a entraîné une rupture dans la façon dont était jusqu'à présent envisagée et implémentée la tâche de résolution des coréférences. Nous pensons que cette rupture impose de remettre en question la conception des mentions en termes de syntagmes maximaux, au moins pour certaines applications dont nous donnons deux exemples. Dans cette perspective, nous proposons une nouvelle formulation de la tâche, basée sur les têtes, accompagnée d'une adaptation du modèle de Lee *et al.* (2017) qui l'implémente.

ABSTRACT

Neural coreference resolution : a head-based approach

End-to-end neural approaches for coreference resolution broke away from the traditional implementation and conception of this task. We argue that, at least for some applications, this breakaway forces us to reconsider the definition of mentions in terms of maximal syntagms : we consider the examples of machine translation and summarization. In this perspective, we propose a new formulation of the task, based on mention heads, and adapt the model of Lee *et al.* (2017) to that end.

MOTS-CLÉS : résolution des coréférences, réseaux de neurones, modèles de bout en bout.

KEYWORDS: coreference resolution, neural networks, end-to-end models.

1 Introduction

La tâche de résolution des coréférences consiste à partitionner des mentions de référents de discours en chaînes de coréférence. Cette définition abstraite ne spécifie pas la nature exacte des mentions de référents de discours, qui demeure à ce jour mystérieuse, et pose la question de la forme linguistique sous laquelle les mentions apparaissent et grâce à laquelle elles peuvent être localisées dans un texte.

C'est une représentation syntaxique des mentions qui a été conventionnellement choisie en TAL : on prend pour définition d'une mention un empan de texte correspondant à un syntagme maximal, c'est-à-dire à la projection maximale d'un des mots de la phrase. Ainsi, la tâche de résolution des coréférences telle qu'elle est pratiquée correspond d'une part à une analyse syntaxique et d'autre part à la résolution des coréférences à proprement parler.

*. Cet article fait suite à un travail réalisé à l'Université Paris Diderot.

Ce choix n'a pas été sans conséquence sur les architectures des systèmes de résolution des coréférences proposées. Il nous apparaît que l'émergence des architectures neuronales de bout en bout — comme celle de Lee *et al.* (2017) — invite aujourd'hui à le rediscuter.

Après avoir évoqué les deux principales architectures de système de résolution des coréférences neuronales (§ 2), nous nous intéresserons à la section 3 à la nature des attentes que peuvent avoir les applications aval par rapport à un système de coréférences (en ne mentionnant brièvement, faute de place, que quelques exemples) avant de proposer une approche de bout en bout basée sur les têtes, dont les premiers résultats que nous rapportons suggèrent qu'elle peut être fructueuse (§ 4).

2 Approches neuronales existantes

2.1 Approche en deux étapes

À la suite de Soon *et al.* (2001), de nombreux modèles statistiques de résolution des coréférences ont été proposés. La majorité d'entre eux utilisait des classeurs ou ordonneurs linéaires pour déterminer des liens de coréférence entre mentions, voire entre une mention et une chaîne de coréférence partielle dans le cas des modèles basés sur les entités.

Ces modèles à l'expressivité limitée impliquaient d'extraire indépendamment les descripteurs pertinents pour résoudre les liens de coréférence. Aussi, les architectures proposées procédaient typiquement à une analyse syntaxique préalable à la résolution des coréférences pour extraire à la fois les mentions et des descripteurs pertinents, au premier rang desquels leurs têtes.

Cette architecture en deux étapes, autour de laquelle a longtemps été conçue la résolution des coréférences, permettait de disposer à la fin du processus à la fois des liens de coréférence entre mentions représentées par des empanaux maximaux, et des représentations linguistiques riches de ces mentions sous forme d'arbres syntaxiques.

Suite aux premiers succès de l'apprentissage de représentations neuronales en TAL, Wiseman *et al.* (2015) ont proposé le premier modèle neuronal de résolution des coréférences, qui était donc capable d'apprendre à calculer des représentations vectorielles des mentions pertinentes pour la tâche, avant de procéder à la traditionnelle étape de résolution des coréférences par ordonnement linéaire.

Ces travaux ont été suivis de plusieurs autres propositions de modèles neuronales (Clark & Manning, 2015, 2016; Wiseman *et al.*, 2016), qui ont permis de très rapidement faire progresser l'état de l'art de plusieurs points.

2.2 Approche de bout en bout

Le modèle introduit par Lee *et al.* (2017) se démarque des systèmes antérieurs en abandonnant l'architecture en deux étapes pour pleinement tirer parti des capacités des réseaux de neurones. Grâce à leur expressivité, ceux-ci sont en effet capables d'apprendre implicitement les descripteurs pertinents, y compris ce qui différencie une mention d'un empan de texte quelconque, rendant de ce fait l'analyse syntaxique implicite.

Concrètement, le modèle considère tous les empanaux à l'intérieur de chaque phrase comme membre

potentiel d'une relation de coréférence. Cela revient à joindre la tâche de détection des mentions et la tâche de résolution des coréférences. Cette approche a permis un gain de performance de 1,5 points CoNLL que les auteurs attribuent à la réduction de la propagation d'erreur typique d'une architecture en pipeline.

Cette approche a cependant comme défaut de produire un résultat beaucoup moins riche que l'architecture en deux étapes : des chaînes de coréférence d'empans maximaux plats, sans aucune information supplémentaire.

Il semble maintenant utile de prendre un peu de recul par rapport à la tâche normalisée : en effet, les performances du modèle de Lee *et al.* (2017) pourraient perdre de leur intérêt si sa sortie se révélait insuffisante pour les applications aval. Nous considérerons brièvement quelques exemples dans la section suivante.

3 Interlude : quelques exemples d'application

De nombreuses tâches qui sont ou pourraient être utilisatrices des résultats d'un résolveur de coréférences ne sont pas intéressées par les bornes des mentions telles que définies dans la tâche standard. On peut en effet distinguer deux principaux cas d'utilisation du résultat d'un résolveur de coréférences :

- on souhaite savoir quelles chaînes de coréférence sont présentes dans chacune des phrases, ou tout au plus dans quelle phrase est situé l'antécédent de telle anaphore ;
- on souhaite connaître les empans des mentions, mais aussi leur structure linguistique interne, par exemple pour transformer le texte.

Dans le premier cas, aucune localisation précise des mentions n'est requise et elles peuvent être identifiées de n'importe quelle manière. Dans le second, les délimitations en termes de syntagmes maximaux s'avèrent souvent insuffisantes, l'information requise pouvant aller jusqu'aux arbres syntaxiques complets des mentions.

Sans prétendre si brièvement couvrir de façon représentative l'ensemble des applications utilisatrices de résolution de coréférences, nous allons maintenant décrire une approche de résumé automatique, et une de traduction automatique qui permettent d'illustrer ces considérations.

3.1 Résumé automatique

Le résumé automatique extractif construit le résumé d'un document par sélection de certaines de ses phrases jugées particulièrement informatives, en général par une fonction de score. Durrett *et al.* (2016) remarquent que, sans précautions particulières, 60 % des phrases choisies contiennent des pronoms orphelins. Pour remédier à ce problème, ils proposent d'ajouter des contraintes d'anaphoricité dans leur système par deux approches complémentaires, toutes deux basées sur un système de résolution des coréférences.

La première méthode s'applique lorsque, pour un pronom donné, une chaîne de coréférence est prédite par le système avec une forte probabilité. Si une mention de l'entité n'a pas encore été incluse dans le résumé, le pronom est remplacé par la première mention de sa chaîne de coréférence. Dans le cas

contraire (deuxième méthode), le système force l'inclusion dans le résumé de contenu supplémentaire, de façon à garantir la clarté de la référence du pronom.

Dans ce dernier cas, les syntagmes maximaux ne jouent aucun rôle et il suffirait amplement, par exemple, d'identifier les différentes mentions par leurs têtes. En ce qui concerne le premier cas, une note de bas de page précise que la première mention de la chaîne de coréférence du pronom n'est en fait pas toujours utilisée telle quelle.

En effet, si la tête de la mention est un nom propre, le pronom n'est remplacé que par la partie de la mention qui correspond au nom propre, plutôt que par le syntagme nominal entier. Mais même dans le cas d'un syntagme nominal standard, le remplacement aveugle d'un pronom par un syntagme maximal pourrait conduire à des incohérences. Les syntagmes maximaux ne sont donc pas, dans ce cas-là non plus, les délimitations les plus pertinentes.

De plus, du fait de la grande hétérogénéité des liens qui unissent deux mentions d'une même chaîne de coréférence, de nombreuses précautions doivent être prises avant de procéder à un remplacement. Durrett *et al.* (2016) précisent qu'ils veillent par exemple à remplacer les adjectifs possessifs par un syntagme possessif.

Cet exemple illustre d'une part, que les besoins en matière de résolution des coréférences sont très variés et que des empan de texte bien délimités ne sont pas toujours nécessaires ; d'autre part, que lorsque des délimitations sont requises, il ne s'agit non seulement pas nécessairement des syntagmes maximaux, mais qu'en plus, beaucoup d'informations sur les mentions sont nécessaires pour pouvoir exploiter efficacement les chaînes de coréférence.

Typiquement, l'analyse syntaxique traditionnellement effectuée en amont de la résolution des coréférences permettrait ici d'identifier les noms propres¹ et de supprimer les propositions relatives des mentions. La gestion des possessifs pourrait quant à elle éventuellement se contenter de quelques heuristiques sur les mots compte tenu de la grande simplicité de l'anglais en la matière.

3.2 Traduction automatique

À la suite de Le Nagard & Koehn (2010), l'utilisation de la résolution des coréférences en traduction automatique a été abondamment décrite. Les principaux modèles de traduction automatique traduisent les phrases du document source une par une indépendamment les unes des autres, ce qui n'est pas sans conséquence sur la cohérence de la traduction.

À titre d'illustration, reprenons l'exemple de Le Nagard & Koehn (2010). *Google Translate* traduisait à l'époque le discours (1a) par (1b). Le pronom *it* est ici mal traduit en l'absence d'information sur le genre de son antécédent.

- (1) a. The window is open. It is black.
b. La fenêtre est ouverte. Il est noir.

Pour éviter ce genre d'incohérences qui, dans des cas moins triviaux, nuisent grandement à la compréhensibilité du texte traduit, Le Nagard & Koehn (2010) proposent de prétraiter les documents en amont du modèle de traduction. Un système de résolution des coréférences est appliqué sur le

1. Dans le très utilisé jeu d'étiquettes morphosyntaxiques du Penn Treebank (Marcus *et al.*, 1993), les noms propres sont clairement marqués par les étiquettes *NNP* et *NNPS*.

document source pour identifier les antécédents des pronoms qui sont remplacés par des pronoms factices porteurs d'information sur le genre que doit avoir leur traduction.

Dans ce cas d'utilisation de la résolution des coréférences, c'est la tête des mentions qui est utile puisque c'est elle qui est porteuse du genre du référent. Un syntagme maximal sans analyse syntaxique pour identifier la tête de la mention s'avérerait de fait inexploitable.

4 Approche de bout en bout basée sur les têtes

4.1 Principe général

D'après ce qui précède, il est parfois souhaitable de disposer des têtes des mentions plutôt que d'empans maximaux. Dans l'approche traditionnelle de la résolution des coréférences par étapes successives d'analyse, cela ne pose aucune difficulté : les mentions sont extraites en amont avec leurs arbres syntaxiques complets dont on peut facilement extraire les têtes. En revanche, dans l'approche de bout en bout basée sur les empans proposée par Lee *et al.* (2017), seuls des empans maximaux plats sont extraits du texte.

Une solution naïve consisterait à appliquer une analyse syntaxique sur le texte après la résolution des coréférences pour extraire les têtes des mentions, mais cela reviendrait à faire deux fois l'analyse syntaxique : une fois implicitement et sous les fortes contraintes de complexité propres à la résolution des coréférences, pour trouver les empans maximaux, et une seconde fois explicitement pour trouver leurs têtes.

Nous avons cependant expérimenté une approche différente : travailler directement sur les têtes, c'est-à-dire choisir les têtes sémantico-syntaxiques plutôt que les syntagmes maximaux pour représenter les mentions. Notons d'ailleurs que c'est le choix qui est à peu près toujours fait dans le cas particulier de la résolution des anaphores événementielles.

Il semble bien sûr impossible de faire une résolution des coréférences de qualité en se basant uniquement sur des mots, mais les architectures neuronales permettent aujourd'hui de contextualiser les mots très efficacement. Par exemple, le modèle de Lee *et al.* (2017) peut être très simplement adapté pour produire des chaînes de coréférence de têtes puisque chaque token est contextualisé par un réseau récurrent bidirectionnel (Schuster & Paliwal, 1997).

La question de l'évaluation d'un système basé sur les têtes est plus simple puisqu'il n'est plus question de décider arbitrairement si des bornes de mentions partiellement correctes, voire approximatives, doivent être sanctionnées plus ou moins sévèrement.² Les métriques usuelles d'évaluation de la résolution des coréférences peuvent être utilisées de la même manière que sur des empans maximaux, avec l'avantage que seule est évaluée la capacité du modèle à détecter la référentialité et à résoudre les relations de coréférences entre les mentions, indépendamment de toute considération syntaxique.

2. Les erreurs dans les bornes des mentions prédites ont été traitées avec une sévérité plus ou moins grande selon l'époque et le corpus utilisé. Lors des campagnes d'évaluation MUC, une prédiction était considérée comme correcte dès lors que la mention prédite contenait la tête annotée et qu'elle ne dépassait pas du syntagme maximal. La plus récente campagne SemEval 2010 s'est montrée un peu plus sévère : 1 point était accordé si la mention prédite coïncidait parfaitement avec celle annotée, mais seulement 0.5 dans le cas où elle contenait sa tête et ne s'étendait pas au-delà des bornes de référence. Enfin, les campagnes d'évaluation CoNLL 2011 et CoNLL 2012 ont introduit la règle en vigueur aujourd'hui : une mention prédite n'est considérée comme correcte que si ses bornes coïncident exactement avec celles de référence.

4.2 Modèle

Pour construire un modèle de résolution des coréférences de bout en bout basé sur les têtes, il suffit de considérer chaque paire de mots comme potentiellement coréférente, plutôt que chaque paire d’empan de phrase. La complexité intrinsèquement quartique de l’approche de Lee *et al.* (2017) qui avait nécessité de nombreuses simplifications pour se ramener à une complexité linéaire (par exemple limiter arbitrairement la longueur des mentions), est de fait ramenée à une complexité intrinsèquement quadratique. La seule hypothèse simplificatrice qui reste nécessaire pour atteindre une complexité linéaire est de limiter la distance à laquelle peuvent se trouver deux mentions coréférentes.

Concrètement, seule la représentation des candidats mentions diffère du modèle basé sur les empan. Dans l’approche proposée par Lee *et al.* (2017), le i -ème empan est représenté par le vecteur g_i , concaténation des vecteurs contextualisés des premier et dernier mots de l’empan, d’un vecteur d’attention à la tête, et d’un vecteur $\phi(i)$ encodant la largeur de l’empan (équation 1).

$$g_i = \left[x_{DEBUT(i)}^*; x_{FIN(i)}^*; \hat{x}_i; \phi(i) \right] \quad (1)$$

Dans notre modèle basé sur les têtes, les candidats mentions sont les mots du texte plutôt que des empan. Aussi, le mot i est simplement représenté par le vecteur contextualisé qui lui correspond, le mécanisme d’attention à la tête étant naturellement abandonné (équation 2).

$$g_i = x_i^* \quad (2)$$

Puisque le vecteur x_i^* est une représentation contextualisée du i -ème mot, vraisemblablement riche en informations syntaxiques et sémantiques, notre approche va cependant bien au-delà de l’heuristique de la même tête proposée par Elsner & Charniak (2010).

4.3 Expériences

Pour évaluer l’intérêt de notre approche, nous comparons notre modèle directement basé sur les têtes à une approche naïve utilisant le modèle basé sur les empan de Lee *et al.* (2017) puis une extraction des têtes à partir des arbres syntaxiques de référence.

Les deux modèles utilisés sont entraînés et évalués sur le corpus CoNLL 2012 (Pradhan *et al.*, 2012), basé sur Ontonotes (Hovy *et al.*, 2006). Ce corpus ne fournissant pas les têtes des mentions, nous les avons déterminées en utilisant une version légèrement modifiée de l’algorithme de recherche de têtes *ModCollinsHeadFinder* (Collins, 1999)³ de la boîte à outils *Stanford CoreNLP* (Manning *et al.*, 2014). Les modifications apportées sont les suivantes :

- (i) les conjonctions de coordination sont considérées comme les têtes des syntagmes nominaux coordonnés pour s’assurer qu’un nom ne peut pas être à la fois la tête du syntagme coordonné et d’un des conjoints ;
- (ii) si aucune règle ne s’applique au constituant, son fils le plus à droite est sélectionné. Cet aménagement est nécessaire pour gérer les étiquettes comme *EMBED* pour lesquelles aucune règle n’existe ;

3. Cet algorithme détermine récursivement la tête d’un constituant à partir des étiquettes syntaxiques de ses enfants.

- (iii) dans le rare cas où une mention n’est pas un constituant syntaxique, le mot le plus à gauche de l’empan est arbitrairement considéré comme sa tête.

Les têtes ainsi obtenues avec la syntaxe de référence sont utilisées dans nos expériences pour :

- (i) construire les exemples d’apprentissage du modèle basé sur les têtes ;
- (ii) produire un corpus d’évaluation annoté en têtes ;
- (iii) transformer les empan maximaux produits par le modèle basé sur les empan, dans l’évaluation de l’approche naïve.

Les scores reportés dans le tableau 1 montrent qu’un important gain de performance se dessine en rappel grâce à la prise en compte de toutes les mentions, sans limite de taille imposée aux syntagmes maximaux, et induit des performances globales supérieures de 1,6 points en termes de F1 CoNLL. Notons également que grâce à la baisse de complexité intrinsèque déjà mentionnée, les temps de calcul et l’empreinte mémoire du modèle basé sur les têtes sont diminués approximativement d’un facteur deux comparé au modèle basé sur les empan de Lee *et al.* (2017).

	MUC			B ³			CEAF _c			CoNLL		
	Prec.	Rec.	F1	Prec.	Rec.	F1	Prec.	Rec.	F1	Prec.	Rec.	F1
Notre approche	78.12	77.84	77.98	68.37	67.3	67.83	63.31	64.32	63.81	69.94	69.80	69.87
Approche naïve	79.62	74.29	76.86	70.08	62.65	66.15	63.66	59.91	61.73	71.12	65.62	68.25

TABLE 1 – Scores de coréférences entre têtes sur le corpus de test CoNLL 2012.

5 Conclusion

La tâche de résolution des coréférences telle qu’on la connaît aujourd’hui n’est qu’une formulation parmi d’autres du problème linguistique sous-jacent. En fonction des applications utilisatrices, des formulations et modèles alternatifs plus appropriés devraient être envisagés.

Nous proposons par exemple une attrayante approche de la tâche, basée sur les têtes, qui pourrait trouver sa place dans plusieurs applications. Cette formulation alternative permet en outre de se focaliser sur la résolution des coréférences à proprement parler en se débarrassant de considérations syntaxiques pas toujours pertinentes.

Comme preuve de concept, nous montrons que lorsqu’il s’agit d’obtenir les têtes des mentions dans l’approche de bout en bout, il est plus efficace de légèrement adapter le modèle proposé par Lee *et al.* (2017) que de post-traiter par analyse syntaxique les empan maximaux qu’il produit.

Remerciements

Ce travail a reçu le soutien du *Labex EFL (Empirical Foundations of Linguistics, ANR-10-LABX-0083)*. Nous remercions Marie Candito, Olga Seminck, Loïc Grobol, Benoît Crabbé pour leurs commentaires critiques et constructifs sur des versions antérieures de cet article, ainsi que les relecteurs de TALN qui ont fourni un retour précieux. Nous remercions enfin Timothee Mickus pour son aide dans la mise en forme de versions précédentes de ce travail.

Références

- CLARK K. & MANNING C. (2015). Entity-Centric Coreference Resolution with Model Stacking. In *Proceedings of the 53rd Annual Meeting of the ACL*.
- CLARK K. & MANNING C. (2016). Deep Reinforcement Learning for Mention-Ranking Coreference Models. In *Proceedings of EMNLP 2016*.
- COLLINS M. (1999). *Head-Driven Statistical Models for Natural Language Parsing*. PhD thesis, University of Pennsylvania.
- DURRETT G., BERG-KIRKPATRICK T. & KLEIN D. (2016). Learning-Based Single-Document Summarization with Compression and Anaphoricity Constraints. In *Proceedings of the 54th Annual Meeting of the ACL*.
- ELSNER M. & CHARNIAK E. (2010). The Same-Head Heuristic for Coreference. In *Proceedings of the ACL 2010 Conference*.
- HOVY E., MARCUS M., PALMER M., RAMSHAW L. & WEISCHEDEL R. (2006). OntoNotes : The 90% Solution. In *Proceedings of the Human Language Technology Conference of the NAACL*.
- LE NAGARD R. & KOEHN P. (2010). Aiding Pronoun Translation with Co-Reference Resolution. In *Proceedings of the Joint 5th Workshop on Statistical Machine Translation*.
- LEE K., HE L., LEWIS M. & ZETTLEMOYER L. (2017). End-to-end Neural Coreference Resolution. In *Proceedings of EMNLP 2017*.
- MANNING C., SURDEANU M., BAUER J., FINKEL J., BETHARD S. & MCCLOSKEY D. (2014). The Stanford CoreNLP natural language processing toolkit. In *Proceedings of 52nd Annual Meeting of the ACL*.
- MARCUS M. P., MARCINKIEWICZ M. A. & SANTORINI B. (1993). Building a large annotated corpus of English : The Penn Treebank. *Computational linguistics*.
- PRADHAN S., MOSCHITTI A., XUE N., URYUPINA O. & ZHANG Y. (2012). CoNLL-2012 shared task : Modeling multilingual unrestricted coreference in OntoNotes. In *Joint Conference on EMNLP and CoNLL-Shared Task*.
- SCHUSTER M. & PALIWAL K. (1997). Bidirectional Recurrent Neural Networks. *Transactions on Signal Processing*.
- SOON W. M., NG H. T. & LIM D. C. Y. (2001). A machine learning approach to coreference resolution of noun phrases. *Computational linguistics*.
- WISEMAN S., RUSH A. & SHIEBER S. (2016). Learning Global Features for Coreference Resolution. In *Proceedings of the 2016 Conference of the NACACL*.
- WISEMAN S., RUSH A., SHIEBER S. & WESTON J. (2015). Learning anaphoricity and antecedent ranking features for coreference resolution. In *Proceedings of the 53rd Annual Meeting of the ACL and the 7th IJCNLP*.