

# Vers une résolution des relations anaphoriques dans la communication électronique médiée

Hani Guenoune<sup>1,2</sup>, Cédric Lopez<sup>2</sup>, Guillaume Tisserant<sup>2</sup>, Mathieu Lafourcade<sup>1</sup>,  
and Melissa Mekaoui<sup>2</sup>

<sup>1</sup> LIRMM / Campus St Priest, 161 Rue Ada, 34090 Montpellier, France

<sup>2</sup> Emvista / Rond-Point Benjamin Franklin, 34000 Montpellier, France

**Résumé** La tâche de résolution des coréférences consiste à identifier toutes les unités linguistiques faisant référence aux mêmes entités, propriétés ou situations du monde réel. Dans la démarche d'acquérir de la connaissance à partir de données textuelles non structurées, il s'agit d'un maillon essentiel aux chaînes d'analyse automatique du langage naturel. L'efficacité des systèmes de résolution automatique dépend directement des ressources utilisées et du caractère rédactionnel du texte auquel ils s'appliquent. Après un sommaire des travaux existants, cet article présentera les enjeux d'une résolution automatique des anaphores dans un contexte de communication électronique en introduisant les verrous susceptibles d'être rencontrés et les pistes envisageables pour les résoudre.

**Mots-clé:** Anaphores · Coréférences · Anaphores abstraites

**Abstract.** The task of coreference resolution applied to written texts consists in finding all lexical parts that refer to the same real world entities, properties or situations. In the aim of extracting knowledge from unstructured textual data, this task plays an essential role within the typical natural language processing pipeline. The efficiency of automated resolution systems depends on the one part, on the resources it relies on, and the redactional nature of the text on the other. After an overview of existing works, this paper will present the issues of dealing with anaphora in the context of electronic communication, the problems one can encounter and consider ways of working around them.

**Keywords:** Coreference · Anaphora · Abstract anaphora.

## Introduction

L'anaphore, phénomène linguistique largement étudié, est abordée selon des angles différents. L'anaphore se définit en rhétorique comme la figure de style consistant à multiplier l'usage d'un syntagme dans le but de renforcer une affirmation ou encore rythmer des propos autour d'une notion centrale [54]. L'anaphore en grammaire est vue comme l'opération imposant une contrainte de dépendance d'interprétation à deux mentions distinctes, dont une première serait

indispensable à la compréhension de la seconde. Elle consiste en une consigne implicite qui dispense de l'interprétation sémantique d'un terme, suggérant plutôt, - pour son interprétation-, une entité lexicale distante. L'identification des référents d'une expression anaphorique est essentielle à la compréhension intégrale d'un discours, permettant notamment à l'allocutaire de connecter les informations relatives aux mêmes entités, la synthèse des caractéristiques citées à différents segments du discours mènent à une définition plus complète des entités dont traite l'énoncé.

Dans la mouvance actuelle de l'intelligence artificielle et du traitement automatique du langage naturel (TALN), plusieurs travaux se consacrent à la question d'identifier les références au sein d'un discours oral ou écrit, à repérer les entités qui sont notées à plus d'un endroit, référencées par plusieurs unités lexicales. L'idée étant de décider si une mention est introduite à l'auditeur, ou est la réapparition d'une notion, propriété ou situation préalablement définie. La tâche de localisation des référents des syntagmes d'un discours occupe une position centrale dans l'architecture des systèmes technologiques visant à tirer profit du langage écrit ou oral. Son accomplissement est nécessaire aux outils de résumé et de traduction automatiques [56; 64], à l'analyse de sentiments [5], ou encore aux systèmes de questions-réponse [7].

La résolution automatique des anaphores se fait aujourd'hui sur la base des rôles grammaticaux des termes, leurs dépendances ainsi que leurs natures sémantiques. Ces exigences renforcent la difficulté de résolution de traitement dans les textes issus de la communication électronique. Dans ce même contexte, la variété des types d'anaphores observées conduit notamment à s'intéresser aux formes d'anaphores les plus répandues dans ce contexte. Nous notons par exemple une utilisation fréquente de l'anaphore abstraite qui consiste à employer des syntagmes (*shell nouns* [28]) qui renvoient vers des entités abstraites représentées syntaxiquement par une, ou un ensemble de phrases.

Cet article commencera par dresser un aperçu des méthodologies employées dans les travaux abordant la résolution automatique des anaphores ainsi que des jeux de données disponibles (*cf.* section 1). Puis, l'article mettra en avant les spécificités de cette tâche en contexte de communication électronique (*cf.* section 2).

## 1 Travaux antérieurs

### 1.1 Terminologie

Dans l'intention de capturer les spécificités des différents schémas de reprise, différentes appellations leur sont assignées. Une distinction émanant de [20] est réalisée par la paire de mots **exophore** et **endophore** dont les préfixes séparent respectivement les caractères externe et interne du contexte dans lequel se définit le référent.

Ainsi, les *endophores* établissent un renvoi vers des éléments internes au discours,

tandis que les *exophores* désignent des entités qui se définissent dans un contexte situationnel, extérieur à l'énonciation et au discours.

**Anaphore** et **cataphore** se distinguent par la localisation relative du terme « source », porteur de la sémantique qui se veut rapportée ou propagée.

L'anaphore admet un terme source *antécédent*, mentionné préalablement dans le discours avant d'être repris. À l'opposée, la cataphore attend l'introduction d'un terme source *conséquent*, apparaissant *a posteriori* de la mention de reprise.

Ce dispositif réussit une simplification des appellations, au moyen de l'acceptation d'une terminologie à deux classes distinctes (exophore/endophore), éliminant dès lors la confusion engendrée par la multiplication de nominations. Le terme *exophore* écarte les variantes équivalentes *déixis* et *référence*, ce qui permet au dernier de maintenir son usage élémentaire et générique, celui d'une opération fondamentale sur laquelle on pourrait asseoir la construction linguistique plus complexe de reprise.

La **référence** représentera ainsi une opération de *désignation* agissant entre l'état lexical d'un terme et un objet concret ou abstrait du monde du discours. Le **renvoi**, en revanche, admet un couple d'opérandes lexicales. Le terme renvoyé est soumis à une interprétation sémantique dépendant du contexte lexical interne.

## 1.2 Typologie.

Bien que l'anaphore se présente dans des formes très variées, les travaux portant sur la résolution de l'anaphore adoptent souvent des typologies élémentaires. Le choix des différentes catégories considérées est imposé d'abord par la langue étudiée, ensuite par les besoins des applications pour lesquelles le système est développé. On distingue également l'anaphore *interphrastique*, dans laquelle les termes liés sont cités dans deux phrases différentes, de l'anaphore *intraphrastique* ou l'antécédent est localisé au sein de la même phrase que le terme anaphorique. Les types le plus souvent considérés en TALN sont les suivants [40] :

- *L'anaphore pronominale* représente le type le plus répandu, et consiste en une reprise par un pronom anaphorique.
- *L'anaphore nominale ou lexicale* désigne le renvoi opéré par un nom commun à un antécédent pouvant être un synonyme ou de sémantique proche.
- *L'anaphore d'instance* représente le cas où l'expression anaphorique n'entretient pas de relation d'identité avec l'antécédent mais désigne une instance de la classe sémantique à laquelle appartient l'antécédent.

## 1.3 Ressources textuelles

L'élaboration de corpus de textes annotés en informations référentielles est nécessaire aux recherches sur la résolution des coréférences. La disponibilité de ces données permet l'apprentissage des descripteurs et des marqueurs qui induisent l'existence de liens anaphoriques. Elle rend également possible l'évaluation des systèmes automatiques de résolution, en testant leur efficacité sur une grande quantité de données.

Au delà de leur finalité applicative, ces collections de données sont précieuses pour les études descriptives du langage naturel. Il s'avère toutefois que la création de ces ressources n'est pas une tâche triviale, de multiples schémas d'annotation ont été proposés sans que la communauté ne puisse converger vers une solution unifiée.

Les corpus *MUC* et *ACE* [21; 13] ont été spécifiquement conçus pour les tâches relatives à l'extraction d'information. Les choix d'annotation ont été pensés de sorte à répondre aux besoins des applications d'extraction d'informations, parfois au détriment des définitions formelles de linguistique. Les modificateurs des syntagmes nominaux n'étaient pas pris en compte, seule la tête d'une phrase nominale était annotée, et les phrases verbales ne figuraient pas au sein des mentions annotées. *ACE* limitait également les types sémantiques des syntagmes nominaux considérés à sept types, pertinents au domaine d'extraction de relations [21].

Les corpus *MATE*, *GNOME* [47], *ARRAU* [48], ainsi que *OntoNotes* [51] visent à couvrir l'étude de la coréférence des points de vues théoriques et pratiques, ils ne se consacrent pas spécifiquement à une tâche en particulier et adoptent des schémas d'annotation qui prennent en compte une typologie des anaphores plus complète. Ils considèrent notamment un ensemble plus large de types grammaticaux des mentions à annoter. *MATE* se penche également sur les phénomènes d'ellipses et propose des annotations pour des relations anaphoriques non identitaires (anaphores possessives, associatives).

*ANCOR* est un corpus français annoté en informations coréférentielles, il consiste en une transcription de données orales. Ses grandes collections de données visent à représenter une variété de langages de domaines différents. Il est constitué de trois corpus distincts, les corpus *Accueil\_UBS*, *OTG* et *ESLO*. Le dernier étant le plus grand des trois, il transcrit des échanges dans le cadre d'interviews socio-linguistiques. [43].

*DEMOCRAT* est la ressource la plus récente, annotée en informations référentielles, son apport consiste à prendre en compte les chaînes de références dans le processus d'annotation [30]. La sélection des textes s'est faite de manière à étudier puis capturer les variations des chaînes de référence en fonction des genres discursifs et des époques. Les textes contenus dans le corpus varient selon plusieurs critères ; époque, type de texte (narratif ou non-narratif), genre textuel.

#### 1.4 Systèmes de résolution automatique

Un certain nombre de travaux donnent un large sommaire des approches automatiques à la résolution des expressions référentielles [59; 49; 45; 44]. Sans prétendre à une quelconque exhaustivité, cette section proposera une vue d'ensemble des recherches qui se sont penchées sur la question du traitement automatique des anaphores.

**Méthodes à base de règles.** Après l'exclusion des mentions non-anaphoriques [62; 10] et étant donné un pronom à résoudre, ces méthodes utilisent des filtres heuristiques dans le but d'écarter les antécédents incompatibles, tandis que les contraintes de préférence calculent un score aux candidats restants afin d'en sélectionner le meilleur.

Les préférences se basent sur des informations de différents niveaux linguistiques, avec un accent sur le niveau morpho-syntaxique [22; 6] et sur la théorie du centrage [18; 25; 61]. Quelques travaux définissent également des règles sémantiques en exploitant des ressources externes [17; 50].

Il s'est manifesté dans cette lignée de travaux symboliques, une tendance à remplacer les systèmes à base de connaissances (knowledge-rich) par des systèmes avec peu de connaissances ou de pré-traitements requis, qui cherchaient à se passer de toute sémantique ou de connaissance de sens commun [29], voire se passer complètement d'analyse morpho-syntaxique préalable [26; 3; 38].

Au delà de la récente disponibilité de grands volumes de données, la difficulté d'une pondération manuelle d'un système avec un grand nombre d'attributs incite au passage aux modèles d'apprentissage qui permettent une gestion plus efficace d'ensembles de descripteurs très grands. Dans [37], les auteurs rapportent que leur modèle d'apprentissage surpassait le meilleur paramétrage de leur modèle à base de règles. [69] relèvent toutefois le manque d'adaptabilité des systèmes statistiques aux différentes formes d'anaphores.

La performance des méthodes statistiques a poussé les systèmes à base de règles à repenser la manière dont l'ensemble des attributs était géré. [19] ordonnent les règles définies, qu'ils appliquent successivement en commençant par les plus importantes. L'idée de définir des séquences de filtres et de règles a été développée dans [52] puis dans [31]. Plus récemment, [32] proposent une approche hybride combinant des classifieurs statistiques à des couches symboliques.

**Méthodes statistiques.** Dans la fin des années 90, le domaine de la résolution d'anaphore en TALN a été marqué par une transition vers les systèmes d'apprentissage automatique. Les premiers modèles implémentaient des algorithmes génétiques [41; 39], ou en encore des réseaux bayésiens [16]. La performance des modèles statistiques s'est graduellement améliorée et finit par surpasser les résultats atteints jusque-là par les systèmes symboliques.

*Les modèles mention-pair* basés sur les mentions procèdent de droite à gauche en comparant la mention à résoudre avec les mentions précédentes, une à une, donnant une prédiction binaire sur l'anaphoricité de la paire.

Les stratégies utilisées variaient entre celles dites *link-first* [55; 58] et *best-first* [46; 67; 4].

La démarche consistant à vérifier les mentions par paires peut générer des chaînes violant la contrainte de transitivité de la relation de coréférence, ce qui aboutit dans un contexte global à des incohérences dans le regroupement des entités.

*Les modèles entity-based* basés sur les entités, profitent des vérifications opérées avec les mentions précédentes [35]. Celles-ci se révèlent intéressantes quand une simple paire de mentions ne suffit pas à se prononcer sur l'égalité de leurs références.

*Les modèles de classement ranking models* répondent à la limitation de la classification binaire qui ne donne pas d'intuition sur la raison du choix d'un antécédent au lieu d'un autre. Cette lacune est adressée en instaurant une compétition qui aboutit à un classement des candidats. [14; 68; 23; 11]

*Les modèles d'apprentissage profond* ont permis d'améliorer les systèmes de classement notamment en ce qui concerne la détection des mentions anaphoriques [33; 8; 9; 65; 66].

## 2 Coréférences dans la communication électronique

La communication électronique prend diverses formes, plus ou moins formelles, variant des messages courts (SMS et plateformes de communication instantanée) aux e-mails professionnels. Ces formes de communication obéissent chacune à un ensemble de règles rédactionnelles nouvelles, qui diffèrent des standards littéraires traditionnels. Dans cette partie, nous aborderons les aspects théoriques et techniques appliqués à la communication électronique.

### 2.1 Aspects théoriques

Les formes de reprise observées dans le langage diffèrent dans une énonciation orale ou écrite et sont imposées par les standards de rédaction du domaine traité. La communication électronique adopte un style qui se rapproche du langage oral, favorisant les ellipses et mettant en avant des types d'anaphores qui peuvent se révéler difficiles à traiter. Nous abordons dans la suite deux phénomènes fréquents dans la communication électronique : les anaphores synthétiques et abstraites ainsi que les ellipses.

*Anaphores synthétiques et abstraites.* Elles font l'objet de diverses études linguistiques qui démontrent la particularité des propriétés des types sémantiques de leurs antécédents, le degré d'abstraction et les propriétés générales des discours dans lesquels elles s'utilisent [2; 63].

Comme dans (1), l'anaphore abstraite est un renvoi à un objet sémantique abstrait dont la représentation syntaxique est une phrase nominale ou verbale [2]. Parmi ces objets sémantiques peuvent figurer les événements, les situations, les propositions, ou encore les faits [1].

*Il y a des problèmes concernant le développement de l'application.  
Cete difficulté doit être surmontée rapidement.* (1)

Compte tenu de la difficulté de la tâche d’annotation d’anaphores abstraites, peu de corpus sont disponibles [12]. Les études les plus anciennes de résolution automatique se sont spécialisées dans le langage oral [15; 57; 42]. Récemment, [24; 34; 53] étudient la résolution des coréférences des évènements, et les anaphores abstraites de manière générale [36; 27; 28].

Par opposition à l’anaphore (pro)nominal, la résolution de l’anaphore abstraite s’avère plus contraignante dans le sens où un certain nombre de marqueurs permettant l’extraction et la sélection des bons candidats deviennent inutilisables. Les attributs d’accord de genre et de nombre entre l’anaphore et l’antécédent ainsi que leur correspondance lexicale fournissent des pointeurs très avantageux dans la discrimination entre les antécédents potentiels. Ces vérifications n’étant évidemment pas applicables à une énonciation phrastique, l’absence de ces indicateurs constitue un inconvénient majeur dans l’ambition d’identifier et de résoudre un lien anaphorique.

*Ellipses.* L’emprunt au style oral conduit à une utilisation fréquente d’ellipses de différents types. On retrouve des ellipses nominales (2) notamment dans l’aposition de propositions contrastives décrivant différentes caractéristiques d’une même notion. Des ellipses du sujet et du verbe sont également récurremment observées (resp (3), (4)).

*Il y a de petites bouteilles, tu veux que je prenne des grandes ?* (2)

*Mathieu passe au dépôt et revient ce soir* (3)

*Mathieu arrivera en voiture et Melissa à pied* (4)

*L’équipe est satisfaite de la v2, nous sommes tous soulagés.* (5)

*Ton bureau est plus grand que le mien.* (6)

Il convient également d’évoquer l’utilisation des *anaphores associatives* (5) qui s’emploient entre des termes qui dépendent l’un de l’autre mais qui ne sont pas coréférents (liés par la relation sémantique de méronymie par exemple). Ou encore les *anaphores d’appartenance (dissociées)* (6) dans lesquelles le terme anaphorique désigne une instance différente de la classe sémantique à laquelle appartient l’antécédent.

## 2.2 Difficultés techniques

Dans cette section, nous listons les points difficiles que nous avons identifiés dans le cadre de la résolution automatique d’anaphores dans la communication électronique.

*Écrits non standards.* Dans la communication électronique, les écrits sont qualifiés de “non standards” ; plusieurs typologies ont été dressées pour classer les phénomènes linguistiques [60]. La fiabilité d’un certain nombre de descripteurs linguistiques est donc diminuée. La mise en place d’un système de normalisation et de correction orthographique peut être bénéfique à la résolution des anaphores.

*Références.* Les communicants peuvent échanger des objets numériques (fichiers textes, vidéos, images, *etc.*). La résolution des références doit analyser le contenu de ces fichiers et/ou de leurs méta-données.

*Support de communication.* Un échange textuel entre deux ou plusieurs communicants est partitionné en plusieurs segments de textes formant un fil de discussion. La portée des renvois est étendue au-delà du simple corps du texte. L'anaphore doit dans ce cas être résolue au sein d'une collection de textes qui n'ont pas de liens apparents.

*Absence potentielle d'antécédents.* L'antécédent d'une mention anaphorique peut être cité dans le texte même, dans le fil de discussion, ou encore dans un échange verbal entre les participants. La présence d'exophores, et l'appui sur le contenu de conversations orales sont des éléments qui contribuent à la complexité de la tâche et nécessitent de repenser les structures des ressources utilisées.

*Absence de corpus annotés.* Nous avons mis en évidence en section 1.3 que, malgré les ressources émergentes visant à contribuer à la résolution automatique des anaphores, aucun jeu de données dans un contexte de communication électronique n'est disponible.

## Conclusion

La résolution automatique des anaphores est une tâche dont la difficulté varie selon les types des anaphores étudiés et le caractère rédactionnel des textes auxquels elle s'applique.

Étant donnée la particularité de la structure des échanges électroniques, il est important de construire un corpus annoté en informations référentielles. Un tel jeu de données serait un point de départ à la résolution des anaphores dans la communication électronique :

1. Sa conception nécessitera une réflexion sur la modélisation de la structure des échanges.
2. Le corpus annoté permettra d'identifier les types d'anaphores les plus fréquents dans le contexte de la communication électronique.
3. Le corpus pourra servir de jeu d'apprentissage, de développement et d'évaluation

Une attention particulière doit être accordée aux anaphores abstraites à antécédents phrastiques qui doivent faire l'objet d'études descriptives plus détaillées. Enfin, des efforts de normalisation des textes doivent également être fournis afin de corriger les erreurs grammaticales et lexicales et permettre une analyse morpho-syntaxique satisfaisante.

## Références

- [1] Amsili, P., Denis, P., Roussarie, L., and Umr, C. P. (2005). Anaphores abstraites en français : représentation formelle.



- [2] Asher, N. (1993). Reference to abstract objects in discourse. In *Studies in linguistics and philosophy*.
- [3] Baldwin, B. (1997). Cogniac : High precision coreference with limited knowledge and linguistic resources.
- [4] Bengtson, E. and Roth, D. (2008). Understanding the value of features for coreference resolution. In *EMNLP*.
- [5] Cambria, E. (2016). Affective computing and sentiment analysis. *IEEE Intelligent Systems*, 31 :102–107.
- [6] Carbonell, J. G. and Brown, R. D. (1988). Anaphora resolution : a multi-strategy approach. In *COLING 1988*.
- [7] Castagnola, L. (2002). Anaphora resolution for question answering.
- [8] Clark, K. and Manning, C. D. (2016a). Deep reinforcement learning for mention-ranking coreference models. In *EMNLP*.
- [9] Clark, K. and Manning, C. D. (2016b). Improving coreference resolution by learning entity-level distributed representations. *ArXiv*, abs/1606.01323.
- [10] Danlos, L. (2006). Ilimp : Outil pour repérer les occurrences du pronom impersonnel il. In *Actes de TALN 2006, Louvain, Belgique*.
- [11] Denis, P. and Baldrige, J. (2008). Specialized models and ranking for coreference resolution. In *EMNLP*.
- [12] Dipper, S. and Zinsmeister, H. (2012). Annotating abstract anaphora. *Language Resources and Evaluation*, 46 :37–52.
- [13] Doddington, G. R., Mitchell, A., Przybocki, M. A., Ramshaw, L. A., Strassel, S., and Weischedel, R. M. (2004). The automatic content extraction (ace) program - tasks, data, and evaluation. In *LREC*.
- [14] Durrett, G. and Klein, D. (2013). Easy victories and uphill battles in coreference resolution. In *EMNLP*.
- [15] Eckert, M. and Strube, M. (2000). Dialogue acts, synchronizing units, and anaphora resolution. *J. Semantics*, 17 :51–89.
- [16] Ge, N., Hale, J., and Charniak, E. (1998). A statistical approach to anaphora resolution. In *VLC@COLING/ACL*.
- [17] Godbert, E. and Benoit, F. (2017). Détection de coréférences de bout en bout en français.
- [18] Grosz, B. J., Joshi, A. K., and Weinstein, S. (1995). Centering : A framework for modeling the local coherence of discourse. *Computational Linguistics*, 21 :203–225.
- [19] Haghighi, A. and Klein, D. (2009). Simple coreference resolution with rich syntactic and semantic features. In *EMNLP*.
- [20] Halliday, M. A. K. and Hasan, R. (1976). Cohesion in english.
- [21] Hirschman, L. and Chinchor, N. (1997). Coreference task definition (v3.0, 13 jul 97).
- [22] Hobbs, J. R. (1978). Resolving pronoun references.
- [23] Iida, R., Inui, K., Takamura, H., and Matsumoto, Y. (2003). Incorporating contextual cues in trainable models for coreference resolution.
- [24] Jauhar, S. K., Guerra, R., Pellicer, E. G., and Recasens, M. (2015). Resolving discourse-deictic pronouns : A two-stage approach to do it. In *\*SEM@NAACL-HLT*.

- [25] Kameyama, M. (1997). Intrasentential centering : A case study. *ArXiv*, cmp-lg/9707005.
- [26] Kennedy, C. and Boguraev, B. (1996). Anaphora for everyone : Pronominal anaphora resolution without a parser. In *COLING*.
- [27] Kolhatkar, V. and Hirst, G. (2014). Resolving shell nouns. In *EMNLP*.
- [28] Kolhatkar, V., Zinsmeister, H., and Hirst, G. (2013). Interpreting anaphoric shell nouns using antecedents of cataphoric shell nouns as training data. In *EMNLP*.
- [29] Lappin, S. and Leass, H. J. (1994). An algorithm for pronominal anaphora resolution. *Computational Linguistics*, 20 :535–561.
- [30] Lattice, LiLPa, ICAR, and IHRIM (2019). Democrat. ORTOLANG (Open Resources and TOols for LANGuage) –www.ortolang.fr.
- [31] Lee, H., Chang, A. X., Peirsman, Y., Chambers, N., Surdeanu, M., and Jurafsky, D. (2013). Deterministic coreference resolution based on entity-centric, precision-ranked rules. *Computational Linguistics*, 39 :885–916.
- [32] Lee, H., Surdeanu, M., and Jurafsky, D. (2017a). A scaffolding approach to coreference resolution integrating statistical and rule-based models. *Natural Language Engineering*, 23 :733–762.
- [33] Lee, K., He, L., Lewis, M., and Zettlemoyer, L. S. (2017b). End-to-end neural coreference resolution. In *EMNLP*.
- [34] Lu, J. and Ng, V. (2016). Event coreference resolution with multi-pass sieves. In *LREC*.
- [35] Luo, X., Ittycheriah, A., Jing, H., Kambhatla, N., and Roukos, S. (2004). A mention-synchronous coreference resolution algorithm based on the bell tree. In *ACL*.
- [36] Marasovic, A., Born, L., Opitz, J., and Frank, A. (2017). A mention-ranking model for abstract anaphora resolution. In *EMNLP*.
- [37] McCarthy, J. F. and Lehnert, W. G. (1995). Using decision trees for coreference resolution. In *IJCAI*.
- [38] Mitkov, R. (1998). Robust pronoun resolution with limited knowledge. In *COLING-ACL*.
- [39] Mitkov, R. (2001). Towards a more consistent and comprehensive evaluation of anaphora resolution algorithms and systems. *Applied Artificial Intelligence*, 15 :253–276.
- [40] Mitkov, R. (2007). Anaphora resolution : The state of the art. In *ACL 2007*.
- [41] Mitkov, R., Evans, R., and Orasan, C. (2002). A new, fully automatic version of mitkov’s knowledge-poor pronoun resolution method. In *CICLing*.
- [42] Müller, M.-C. (2008). Fully automatic resolution of ‘it’, ‘this’, and ‘that’ in unrestricted multi-party dialog.
- [43] Muzerelle, J., Lefeuvre, A., Schang, E., Antoine, J.-Y., Pelletier, A., Maurel, D., Eshkol-Taravella, I., and Villaneau, J. (2014). Ancor\_centre, a large free spoken french coreference corpus : description of the resource and reliability measures. In *LREC*.
- [44] Ng, V. (2003). Machine learning for coreference resolution : Recent successes and future challenges.

- [45] Ng, V. (2010). Supervised noun phrase coreference research : The first fifteen years. In *ACL*.
- [46] Ng, V. and Gardent, C. (2002). Improving machine learning approaches to coreference resolution. In *ACL*.
- [47] Poesio, M. (2004). The mate/gnome proposals for anaphoric annotation, revisited. In *SIGDIAL Workshop*.
- [48] Poesio, M. and Artstein, R. (2008). Anaphoric annotation in the arrau corpus. In *LREC*.
- [49] Poesio, M., Ponzetto, S., and Versley, Y. (2011). Computational models of anaphora resolution : A survey.
- [50] Ponzetto, S. P. and Strube, M. (2006). Exploiting semantic role labeling, wordnet and wikipedia for coreference resolution. In *HLT-NAACL*.
- [51] Pradhan, S., Ramshaw, L. A., Weischedel, R. M., MacBride, J., and Micciulla, L. (2007). Unrestricted coreference : Identifying entities and events in ontonotes. *International Conference on Semantic Computing (ICSC 2007)*, pages 446–453.
- [52] Raghunathan, K., Lee, H., Rangarajan, S., Chambers, N., Surdeanu, M., Jurafsky, D., and Manning, C. D. (2010). A multi-pass sieve for coreference resolution. In *EMNLP*.
- [53] Rajagopal, D., Hovy, E. H., and Mitamura, T. (2016). Unsupervised event coreference for abstract words.
- [54] Reichler-Béguelin, M.-J. (1988). Anaphore, cataphore et mémoire discursive. *Pratiques*, 57(1) :15–43.
- [55] Soon, W. M., Ng, H. T., and Lim, C. Y. (2001). A machine learning approach to coreference resolution of noun phrases. *Computational Linguistics*, 27 :521–544.
- [56] Steinberger, J., Poesio, M., Kabadjov, M. A., and Jezek, K. (2007). Two uses of anaphora resolution in summarization. *Inf. Process. Manage.*, 43 :1663–1680.
- [57] Strube, M. and Müller, C. (2003). A machine learning approach to pronoun resolution in spoken dialogue. In *ACL*.
- [58] Strube, M., Rapp, S., and Müller, C. (2002). The influence of minimum edit distance on reference resolution. In *EMNLP*.
- [59] Sukthanker, R., Poria, S., Cambria, E., and Thirunavukarasu, R. (2018). Anaphora and coreference resolution : A review. *ArXiv*, abs/1805.11824.
- [60] Tarrade, L., Lopez, C., Panckhurst, R., and Antoniadis, G. (2017). Typologies pour l’annotation de textes non standard en français. In *TALN*, pages 118–125.
- [61] Tetreault, J. R. (2001). A corpus-based evaluation of centering and pronoun resolution. *Computational Linguistics*, 27 :507–520.
- [62] Uryupina, O., Kabadjov, M. A., and Poesio, M. (2016). Detecting non-reference and non-anaphoricity. In *Anaphora Resolution - Algorithms, Resources, and Applications*.
- [63] Webber, B. L. (1991). Structure and ostension in the interpretation of discourse deixis. *ArXiv*, cmp-lg/9708003.

- [64] Werlen, L. M. and Popescu-Belis, A. (2017). Using coreference links to improve spanish-to-english machine translation. pages 30–40.
- [65] Wiseman, S., Rush, A. M., and Shieber, S. M. (2016). Learning global features for coreference resolution. In *HLT-NAACL*.
- [66] Wiseman, S., Rush, A. M., Shieber, S. M., and Weston, J. (2015). Learning anaphoricity and antecedent ranking features for coreference resolution. In *ACL*.
- [67] Yang, X. and Su, J. (2007). Coreference resolution using semantic relatedness information from automatically discovered patterns. In *ACL*.
- [68] Yang, X., Zhou, G., Su, J., and Tan, C. L. (2003). Coreference resolution using competition learning approach. In *ACL*.
- [69] Zeldes, A. and Zhang, S. (2016). When annotation schemes change rules help : A configurable approach to coreference resolution beyond ontonotes. In *CORBON@HLT-NAACL*.