

La résolution d'anaphores au-delà de la frontière de la phrase

Luka Nerima, Eric Wehrli

Laboratoire d'analyse et de technologie du langage (LATL)

CUI - Université de Genève, 7 route de Drize, CH-1227 Carouge, Suisse

luka.nerima@unige.ch, eric.wehrli@unige.ch

RÉSUMÉ

Cette démonstration présente une extension de nos outils d'analyse syntaxique et d'étiquetage morphosyntaxique qui prend en compte la résolution d'anaphores pronominales non seulement à l'intérieur d'une phrase, mais également si l'antécédent se trouve dans la phrase précédente. Autant l'analyseur que l'étiqueteur effectuant une analyse syntaxique complète des phrases, ces outils affichent également les fonctions grammaticales des constituants (sujet, objet direct, etc.) et les arguments des verbes. Une version de cette démonstration est disponible sur le Web.

ABSTRACT

The Anaphora Resolution Beyond Sentence Boundary

This demo presents an extension of our syntactic analysis and morphosyntactic tagging tools that takes into account the resolution of pronominal anaphors not only within a single sentence, but also if the antecedent is in the previous sentence. A version of this demo is available on the Web.

MOTS-CLES : Résolution d'anaphores, analyseur syntaxique, étiquetage morphosyntaxique.

KEYWORDS : Anaphora resolution, natural language parsing, POS tagging.

1 Introduction

Malgré les avancées spectaculaires sur les différents fronts des applications de TAL ces 5 dernières années, la résolution d'anaphores (RA) n'a pas progressé de manière significative. Les exemples de tâches où cette résolution joue un rôle important sont pourtant nombreux : traduction automatique, résumés automatiques, reconnaissance d'entités nommées, extraction et identification d'expressions polylexicales, etc. (voir p. ex. Mitkov, 2002, Sukthanker 2018). L'étiqueteur morphosyntaxique que nous montrons ici prend en charge une résolution d'anaphore particulière, celle des antécédents des pronoms personnels de la troisième personne. À noter aussi que, dans leur version actuelle, nos outils d'analyse ne traitent pas les cataphores.

2 La méthode

Dans le cadre de la grammaire générative, la théorie du liage de Chomsky (1981) et les implémentations informatiques de Hobbs (1978) et de Lappin et Leass (1994) sont sans doute les contributions les plus significatives à la RA. La théorie de la liaison de Chomsky n'est pas destinée à

être une méthode de RA en soi, mais un ensemble de contraintes sur la distribution des pronoms réfléchis et/ou réciproques (appelés anaphores) et des pronoms référentiels (appelés pronoms) qui sont extrêmement utiles pour filtrer l'ensemble des antécédents potentiels.

En bref, la théorie du liage stipule (i) que les pronoms réfléchis et/ou réciproques doivent être liés dans leur catégorie minimale de gouvernance, et (ii) que les pronoms référentiels ne peuvent pas être liés dans leur catégorie minimale de gouvernance, où "lié" signifie gouverné et co-indexé. En résumé, nous pouvons définir la catégorie minimale de gouvernance d'un élément comme le constituant minimal qui l'inclut. Il découle des principes du liage (i) et (ii) que les anaphores et les pronoms (pour reprendre la terminologie de Chomsky) ont une distribution complémentaire.

3 La démonstration

Dans les grandes lignes, la mise en œuvre de la RA suit l'algorithme de Lappin et Leass (1994), adapté aux spécificités des représentations grammaticales de notre analyseur. Appliquée à l'exemple de la Figure 1, la première tâche vérifie que le pronom *il* (à la ligne 3 de la deuxième phrase) n'est pas impersonnel (comme p. ex. dans *il pleut*). Pour cela, elle s'appuie sur des caractéristiques lexicales (p. ex. *pouvoir* est un verbe non météorologique) et/ou syntaxiques.

L'étape suivante consiste à déterminer quels sont les antécédents possibles d'un pronom référentiel. Le principe (ii) du liage nous dit qu'il ne peut se trouver dans la proposition qui contient le pronom, la recherche s'effectue donc dans les propositions ou phrases précédentes et concerne tous les syntagmes nominaux qui s'accordent en nombre et en genre avec le pronom.

En cas d'ambiguïté (plusieurs antécédents possibles), on recourt à des heuristiques basées sur la fonction grammaticale de l'antécédent (p. ex. le sujet est prioritaire par rapport à l'objet direct) et des éventuelles restrictions sélectionnelles associées à la fonction du pronom (on s'appuie sur les traits sélectionnels sémantiques des arguments du verbe, p. ex. animé, inanimé, humain, etc.). La Figure 1 montre la sortie produite par l'étiqueteur morphosyntaxique sur la séquence de phrases « *Jean a beaucoup de problèmes. Tellement qu'il ne pourra jamais les résoudre tous* ».

1	Jean	NOM-PRO-SIN-MAS	Jean	SU	
2	a	VER-IND-PRE-3-SIN	avoir		SU:Jean 1 DO:problèmes 5
3	beaucoup	DET-QUA-INN-ING	beaucoup	DO	
4	de	PRE	de		
5	problèmes	NOM-COM-PLU-MAS	problème		
6	.				
1	Tellement	ADV	tellement		
2	qu'	ADV	que		
3	il	PRO-PER-3-SIN-MAS	il	SU	Jean
4	ne	ADV	ne		
5	pourra	VER-IND-FUT-3-SIN	pouvoir		SU:il 3 SO:résoudre 8
6	jamais	ADV	jamais		
7	les	PRO-CLI-PLU-ING	le		
8	résoudre	VER-INF	résoudre		SU:il 3 DO:les résoudre les problèmes
9	tous	PRO-IND-PLU-MAS	tous		
10	.				

Figure 1 : Sortie de l'étiqueteur produite sur deux phrases, avec résolution d'anaphores

Pour chaque mot de la phrase, l'étiqueteur morphosyntaxique affiche les informations suivantes, réparties en 7 colonnes :

1. le numéro d'ordre du mot dans la phrase (utilisé pour référence),
2. la forme orthographique,
3. l'étiquette morphosyntaxique,
4. le lexème,
5. la fonction grammaticale (*SU* : sujet, *DO* : objet direct, *SO* : complément phrase, etc.)
6. les arguments (ou valence) pour les verbes, l'antécédent pour les pronoms anaphoriques,
7. la collocation identifiée (*résoudre les problèmes* dans l'exemple de la Figure 1).

L'analyseur conserve les syntagmes nominaux avec leur fonction grammaticale dans une structure de données ad hoc afin de pouvoir effectuer la RA également dans la(les) phrase(s) précédente(s). Pour une description complète de la résolution d'anaphores implémentée dans notre système, nous invitons le lecteur à consulter l'article de Wehrli, Seretan et Nerima (2020).

Références

CHOMSKY N. (1981). *Lectures on Government and Binding*, Foris Publications.

HOBBS J. (1978). Resolving pronoun references, *Lingua* 44:311-338. DOI : [10.1016/0024-3841\(78\)90006-2](https://doi.org/10.1016/0024-3841(78)90006-2)

LAPPIN S., LEASS, H.J. (1994). An Algorithm for Pronominal Anaphora Resolution, *Computational Linguistics*, 20:4, 535-561. Anthologie ACL : [J94-4002](https://doi.org/10.1016/j.cll.1994.04.002)

MITKOV R. (2014). *Anaphora resolution*. Routledge.

SUKTHANKER R., PORIA S., CAMBRIA E., THIRUNAVUKARASU R. (2018). Anaphora and Coreference Resolution: A Review. Consulté le 12.2.2020. DOI : [10.1016/j.inffus.2020.01.010](https://doi.org/10.1016/j.inffus.2020.01.010)

WEHRLI E., SERETAN, V., NERIMA, L. (2020). Verbal collocations and pronominalisation. In G. C. Pastor & J-P Colson, Éd., *Computational Phraseology, IVITRA Research in Linguistics and Literature*, 20, p. 208–224. John Benjamins. DOI : [10.1075/ivitra.24.11weh](https://doi.org/10.1075/ivitra.24.11weh)

Adresse URL de la démonstration

FIPS : <http://latlapps.unige.ch/Parser/>, application « FullTagger », consulté le 24.04.2020